Simple Linear Regression

Fundamentals

Rahul Singh
rsingh@arrsingh.com

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

in other words…

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

$$y = f(x)$$

in other words…

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

in other words…

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

$$y = f(x)$$

$\boxed{x \text{ is the independent variable}}$

What is Simple Linear Regression

re·gres·sion

noun

A statistical method used to **predict** the relationship between a dependent variable and one or more independent variables

$$y = f(x)$$

in other words…

if we see some data (x, y) we can use linear regression to **predict** the y values for other values of x

x is the independent variable

y is the dependent variable

Lets take a simple example…

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

A simple example…

A car is traveling at a constant speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

Question: Can we **predict** how far the car will have traveled in 4.7 hours?

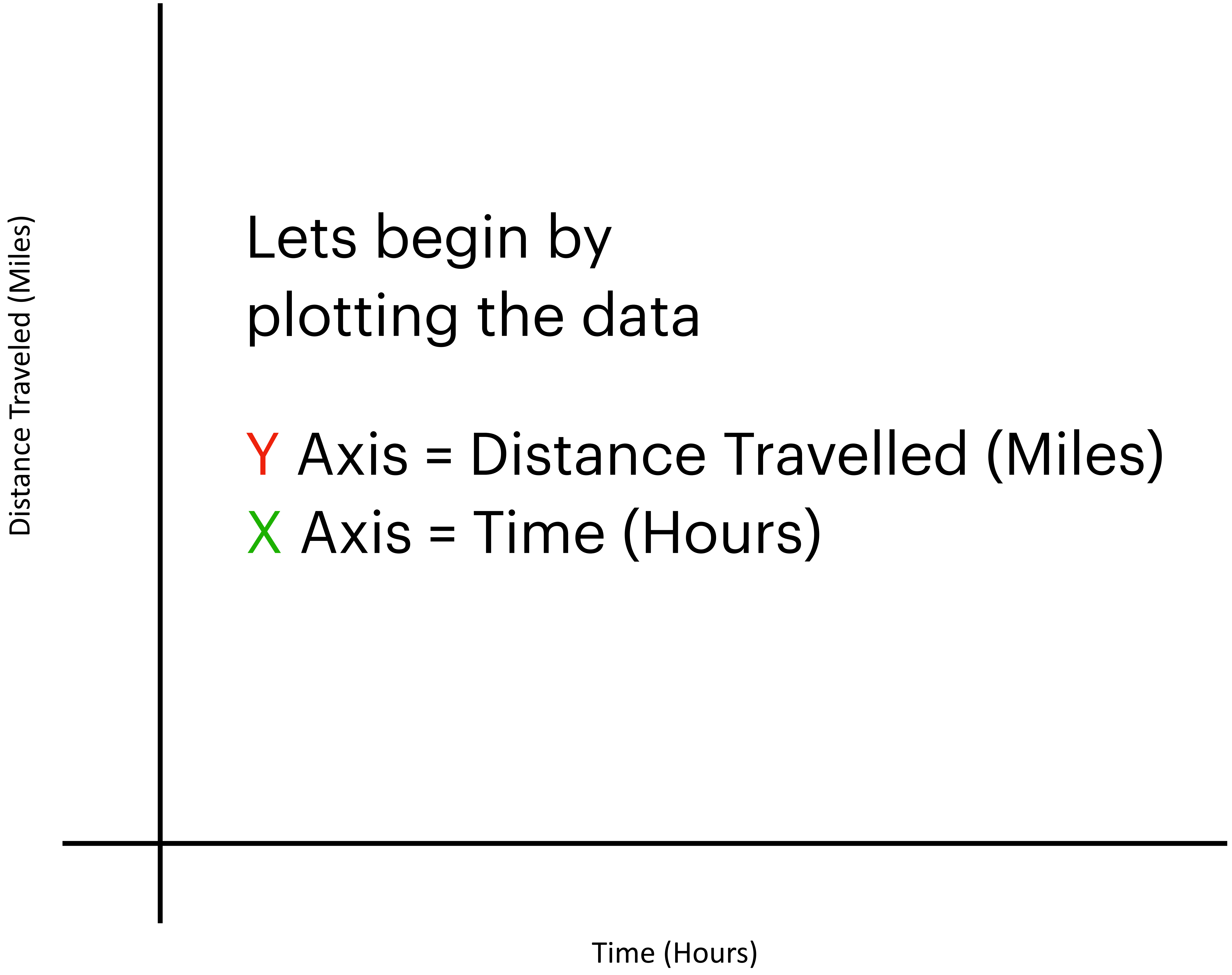
Simple Linear Regression

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

Question: Can we **predict** how far the car will have traveled in 4.7 hours?



Lets begin by plotting the data

Y Axis = Distance Travelled (Miles)

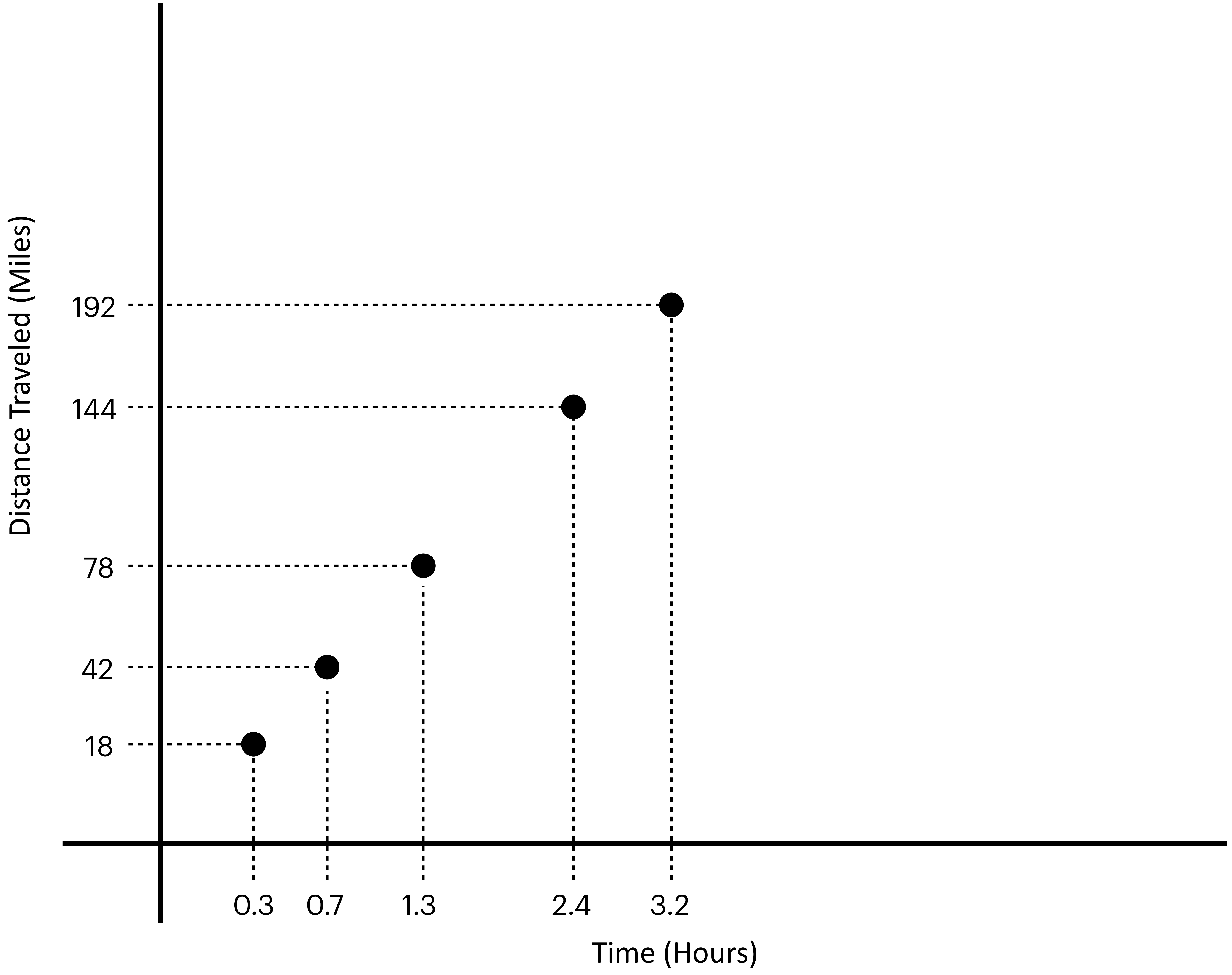
X Axis = Time (Hours)

Simple Linear Regression

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192



Question: Can we **predict** how far the car will have traveled in 4.7 hours?

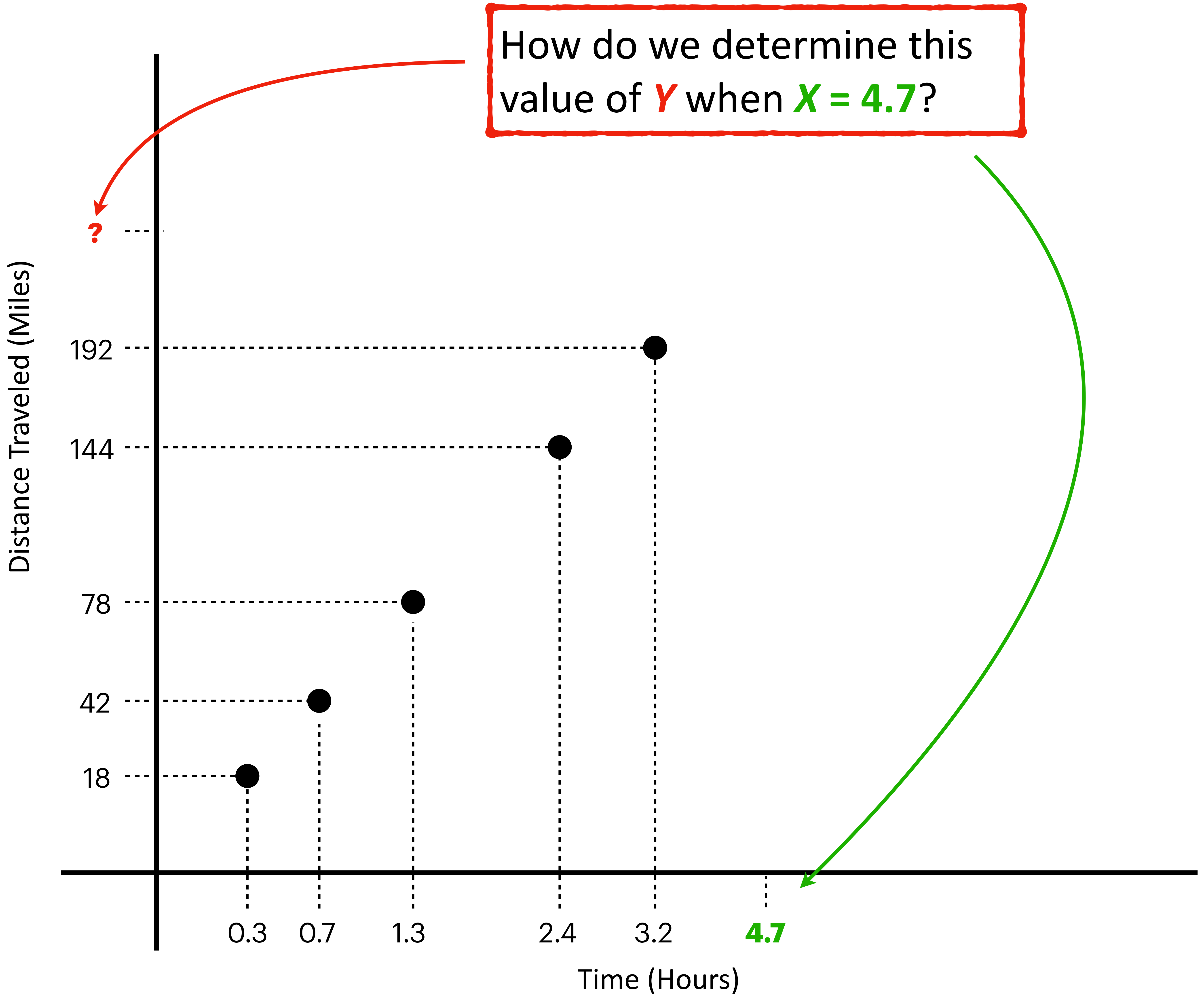
Simple Linear Regression

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

Question: Can we **predict** how far the car will have traveled in 4.7 hours?



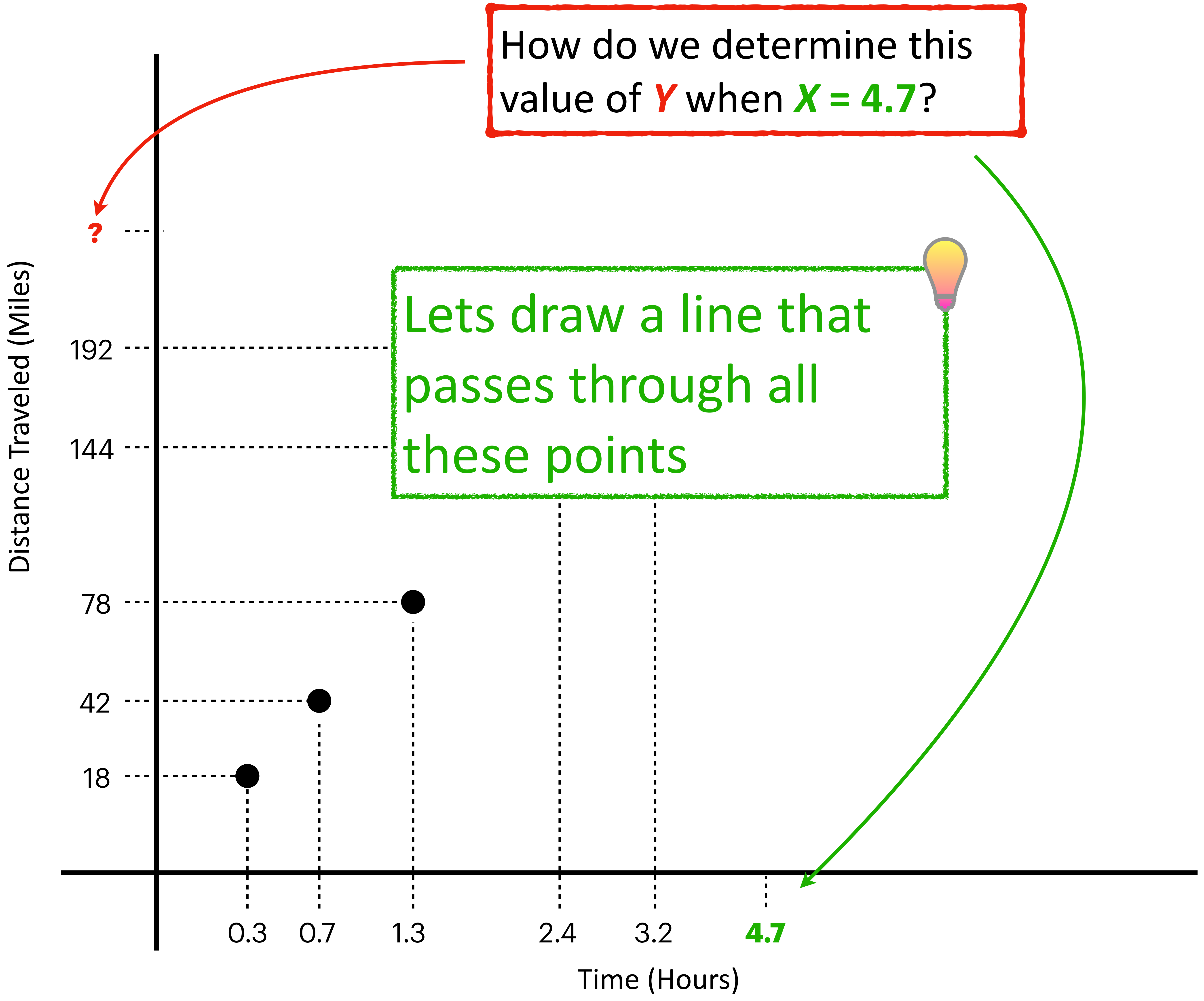
Simple Linear Regression

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

Question: Can we **predict** how far the car will have traveled in 4.7 hours?

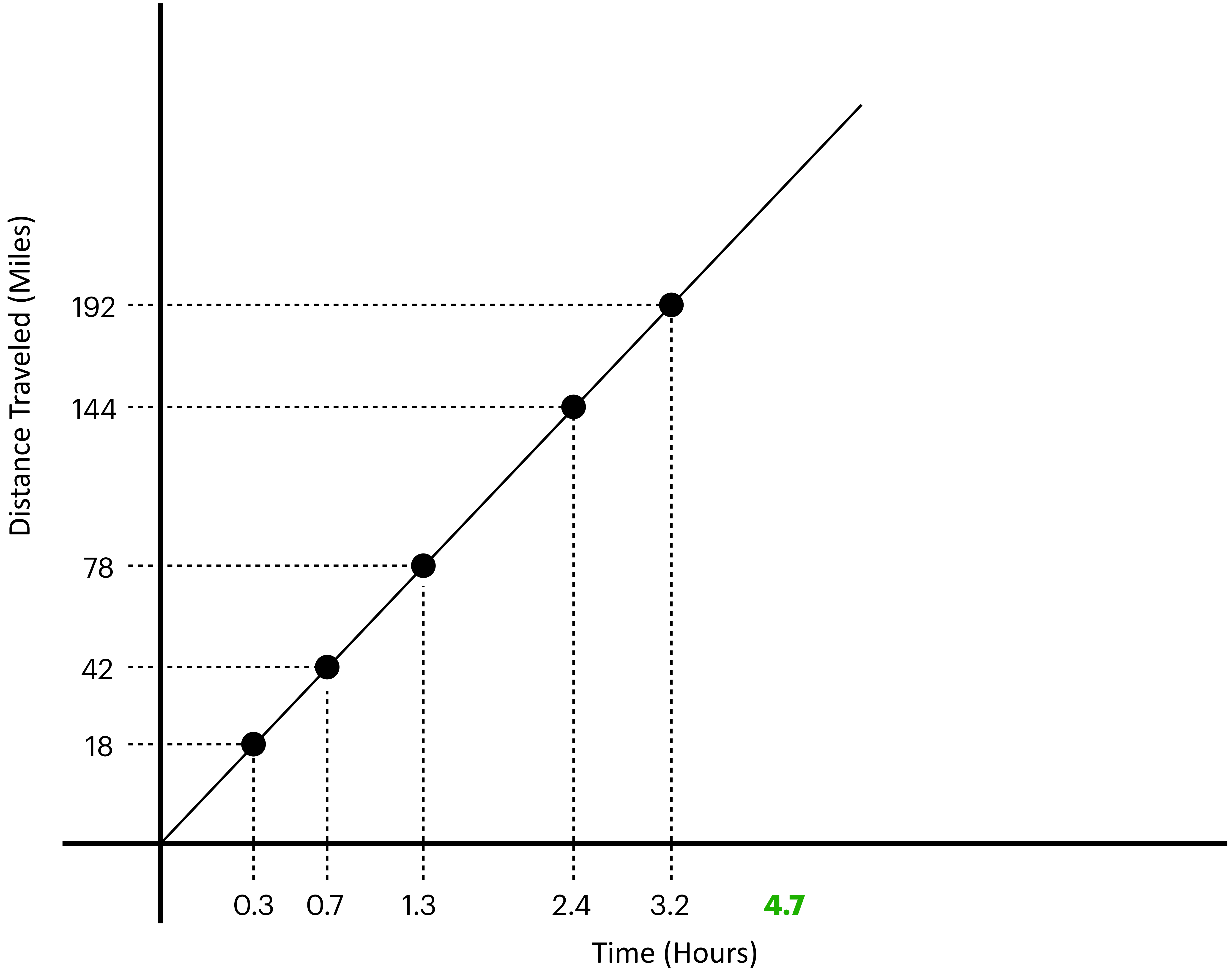


Simple Linear Regression

A simple example…

A car is traveling at a constant speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192



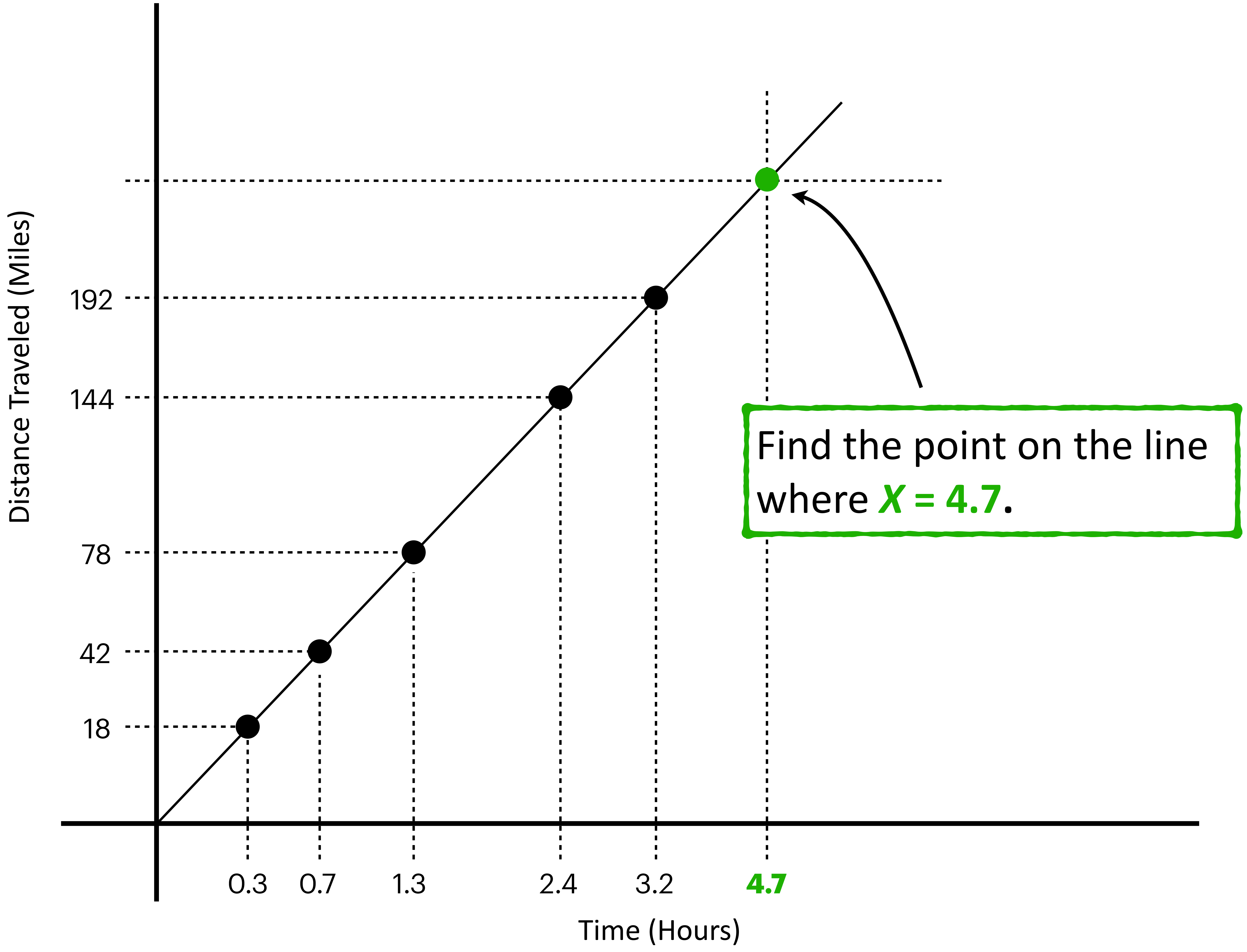
Question: Can we **predict** how far the car will have traveled in 4.7 hours?

A simple example…

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

Question: Can we **predict** how far the car will have traveled in 4.7 hours?



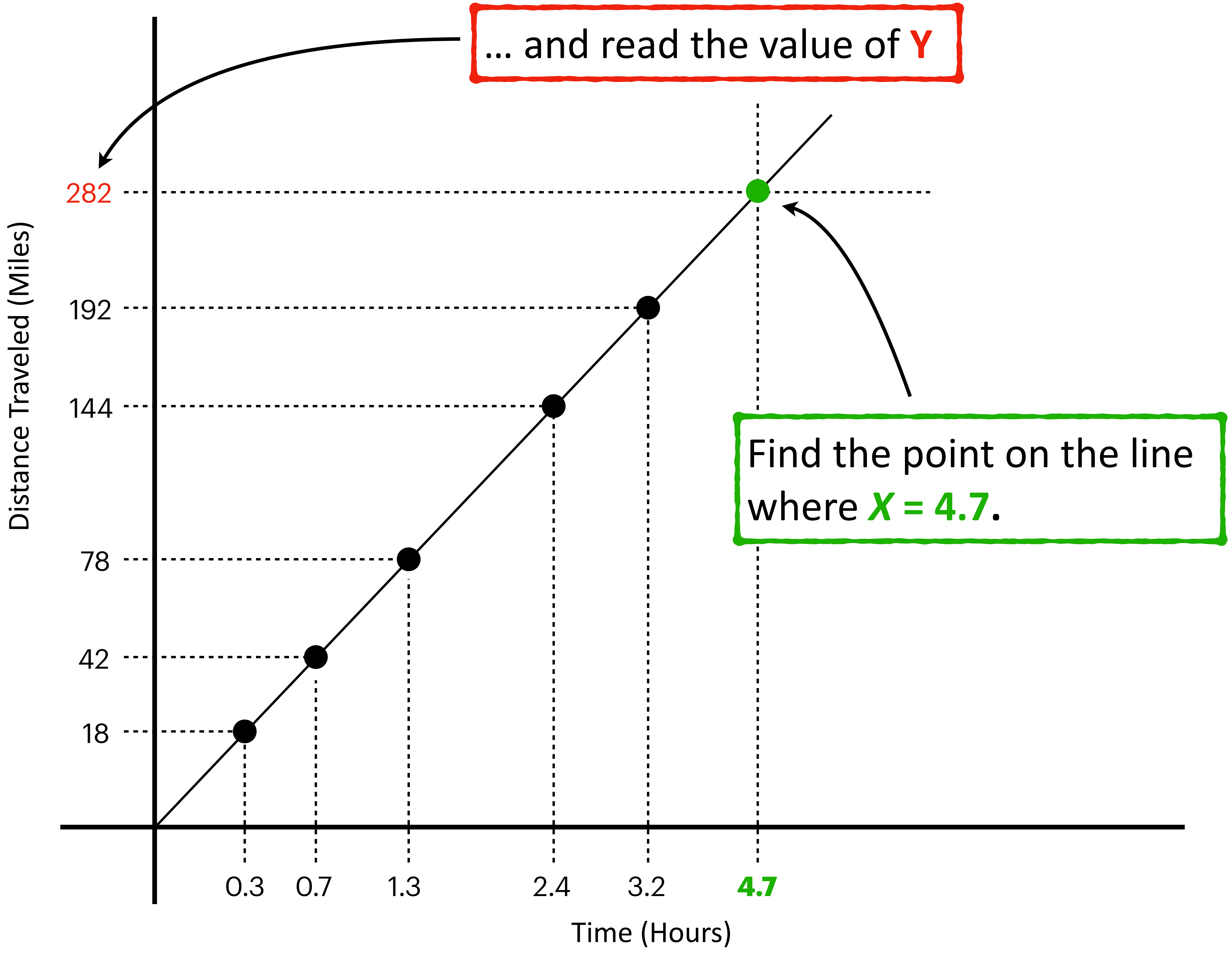
Simple Linear Regression

A simple example...

A car is traveling at a <u>constant</u> speed. We observe the distance travelled by the car at various times during its journey.

Time (Hours)	Distance Traveled (Miles)
0.3	18
0.7	42
1.3	78
2.4	144
3.2	192

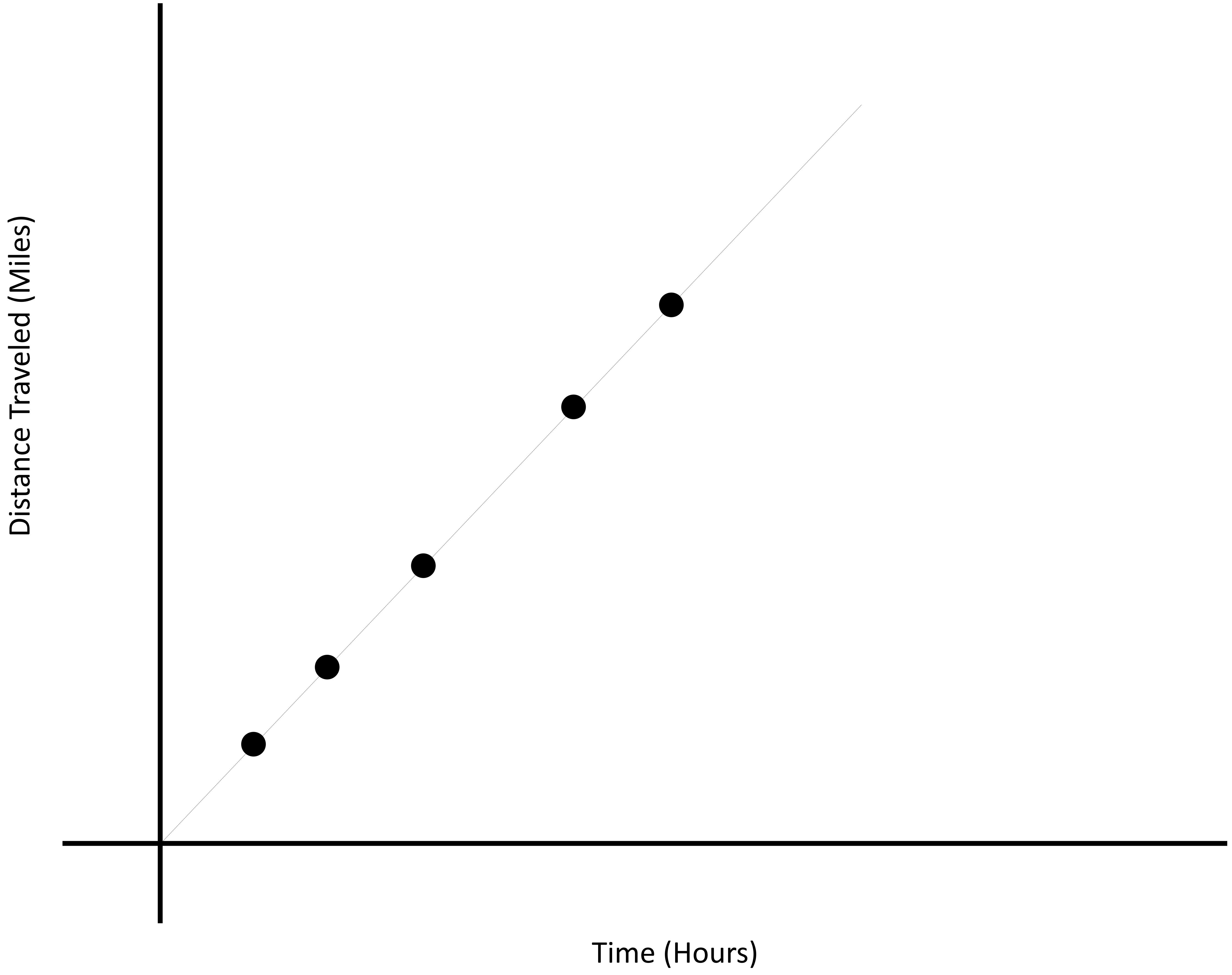
Question: Can we **predict** how far the car will have traveled in 4.7 hours?



Simple Linear Regression

A few things to note…

- All the data points line up perfectly
- The slope of the line (i.e. speed of the car) is easy to determine:

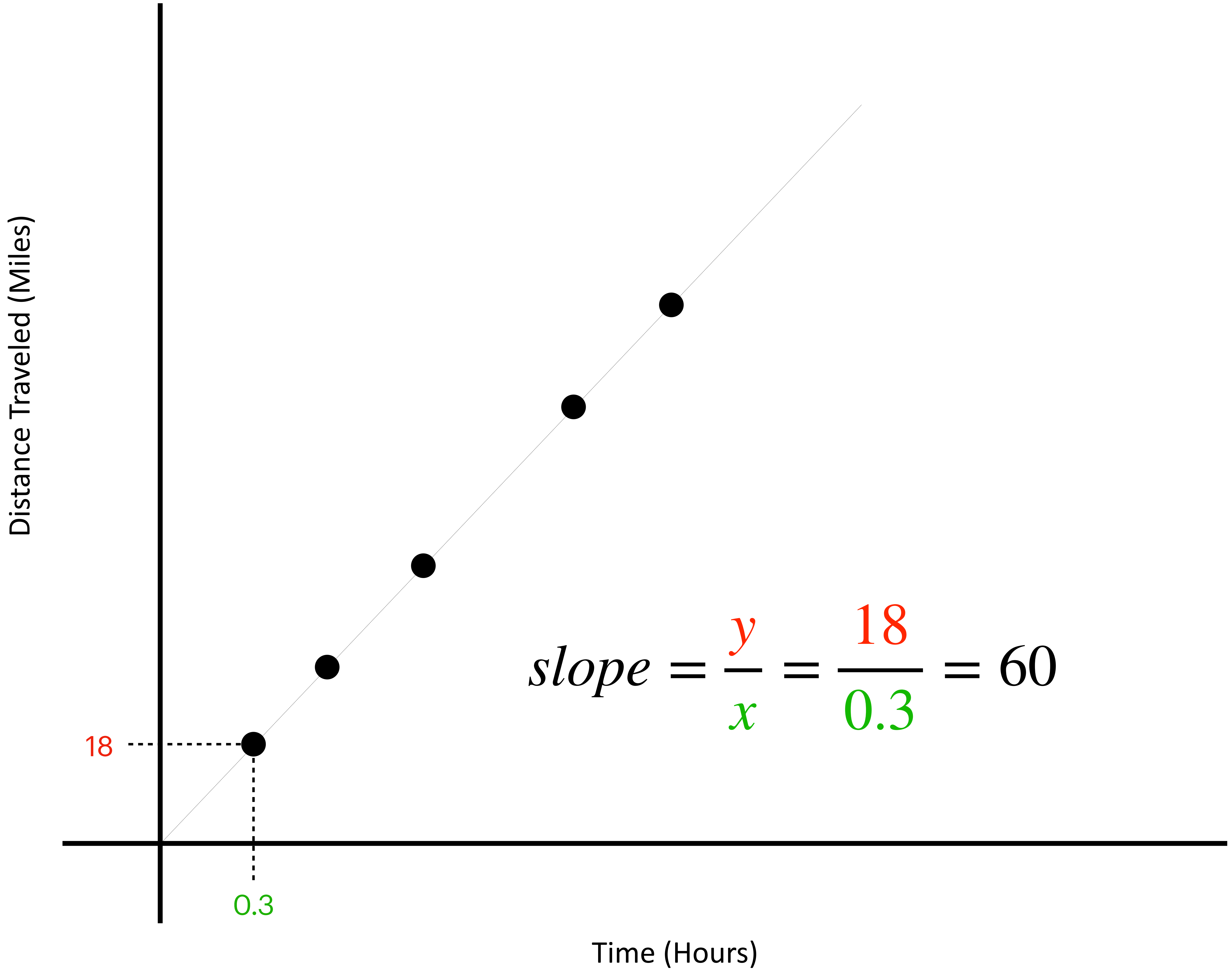


Simple Linear Regression

The slope of the line (i.e. speed of the car) is easy to determine:

Slope at $x = 0.3$, $y = 18$

Time (Hours)	Distance Traveled (Miles)	Speed (mph)
0.3	18	18 / 0.3 = 60



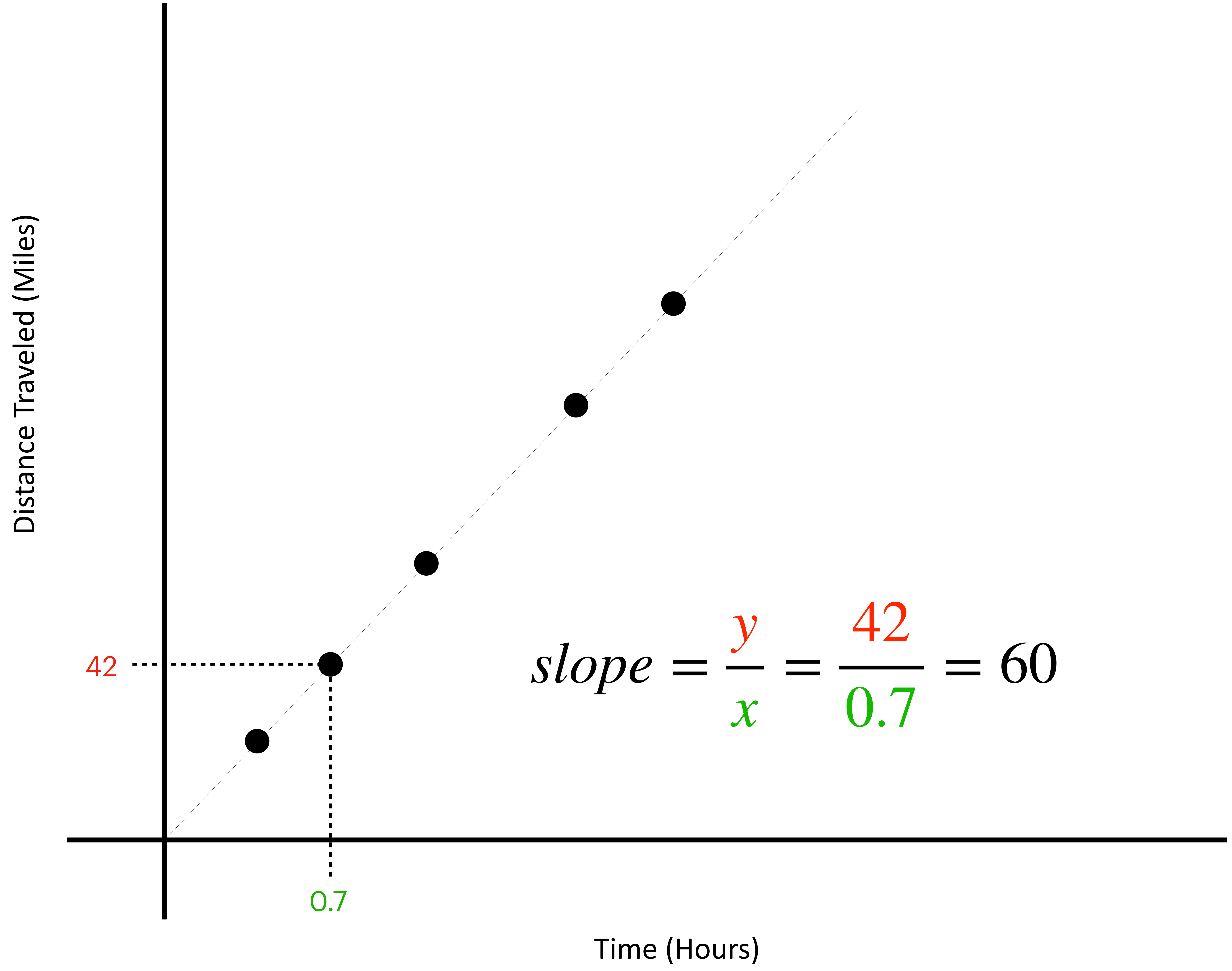
$$slope = \frac{y}{x} = \frac{18}{0.3} = 60$$

Simple Linear Regression

The slope of the line (i.e. speed of the car) is easy to determine:

Slope at $x = 0.7$, $y = 42$

Time (Hours)	Distance Traveled (Miles)	Speed (mph)
0.3	18	18 / 0.3 = 60
0.7	42	42 / 0.7 = 60



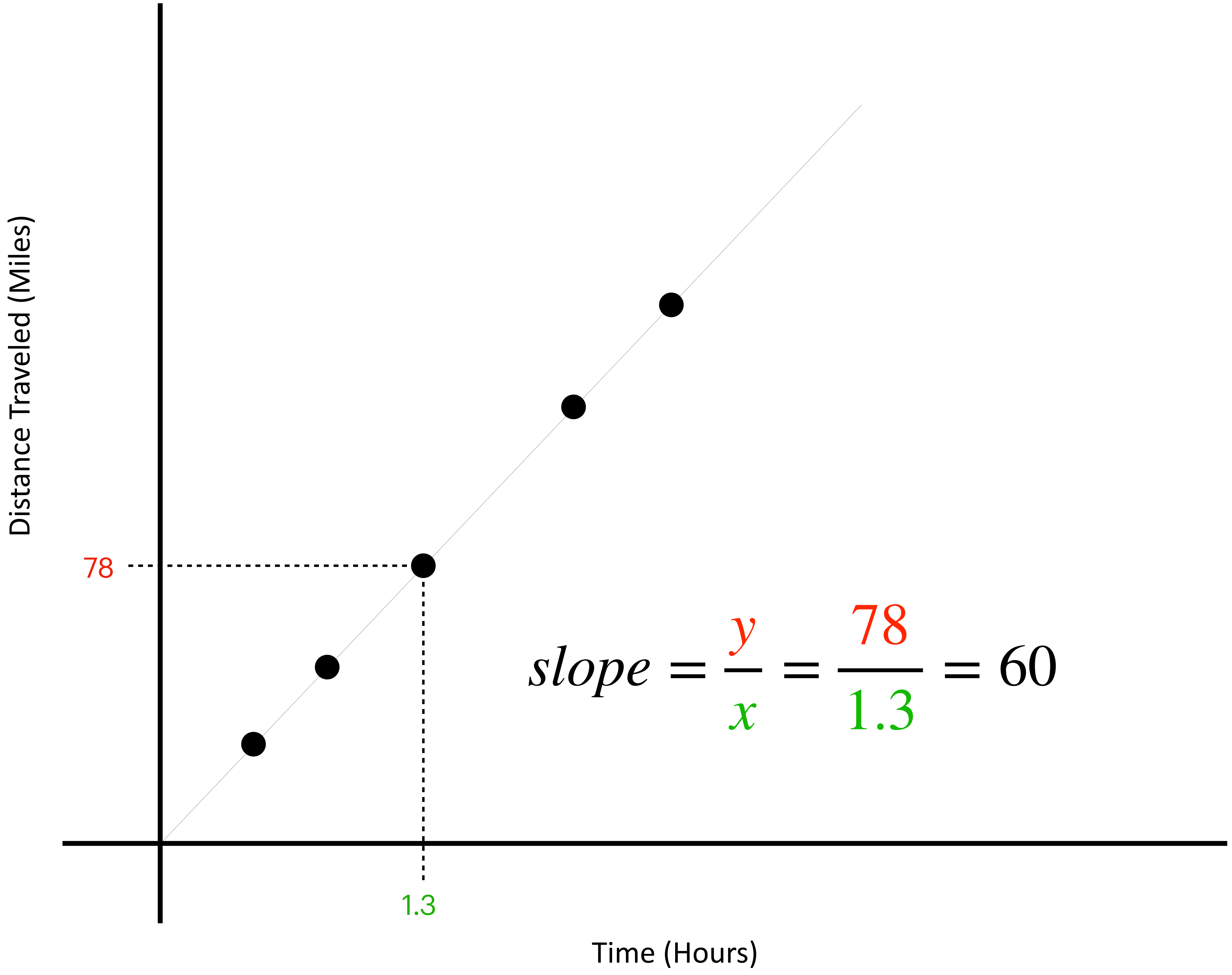
$$slope = \frac{y}{x} = \frac{42}{0.7} = 60$$

Simple Linear Regression

The slope of the line (i.e. speed of the car) is easy to determine:

Slope at $x = 1.3$, $y = 78$

Time (Hours)	Distance Traveled (Miles)	Speed (mph)
0.3	18	18 / 0.3 = 60
0.7	42	42 / 0.7 = 60
1.3	78	78 / 1.3 = 60



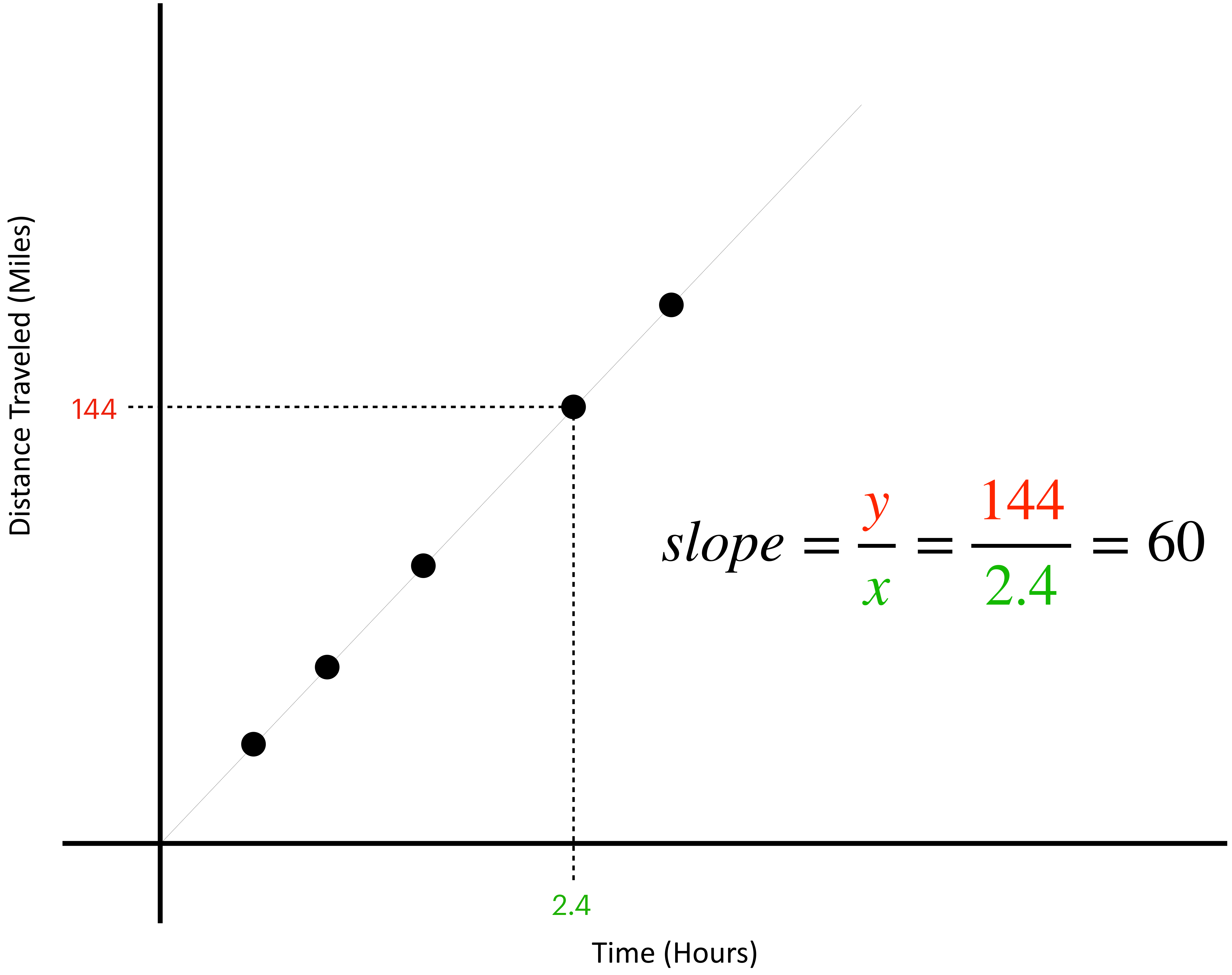
$$slope = \frac{y}{x} = \frac{78}{1.3} = 60$$

Simple Linear Regression

The slope of the line (i.e. speed of the car) is easy to determine:

Slope at $x = 2.4$, $y = 144$

Time (Hours)	Distance Traveled (Miles)	Speed (mph)
0.3	18	18 / 0.3 = 60
0.7	42	42 / 0.7 = 60
1.3	78	78 / 1.3 = 60
2.4	144	144 / 2.4 = 60

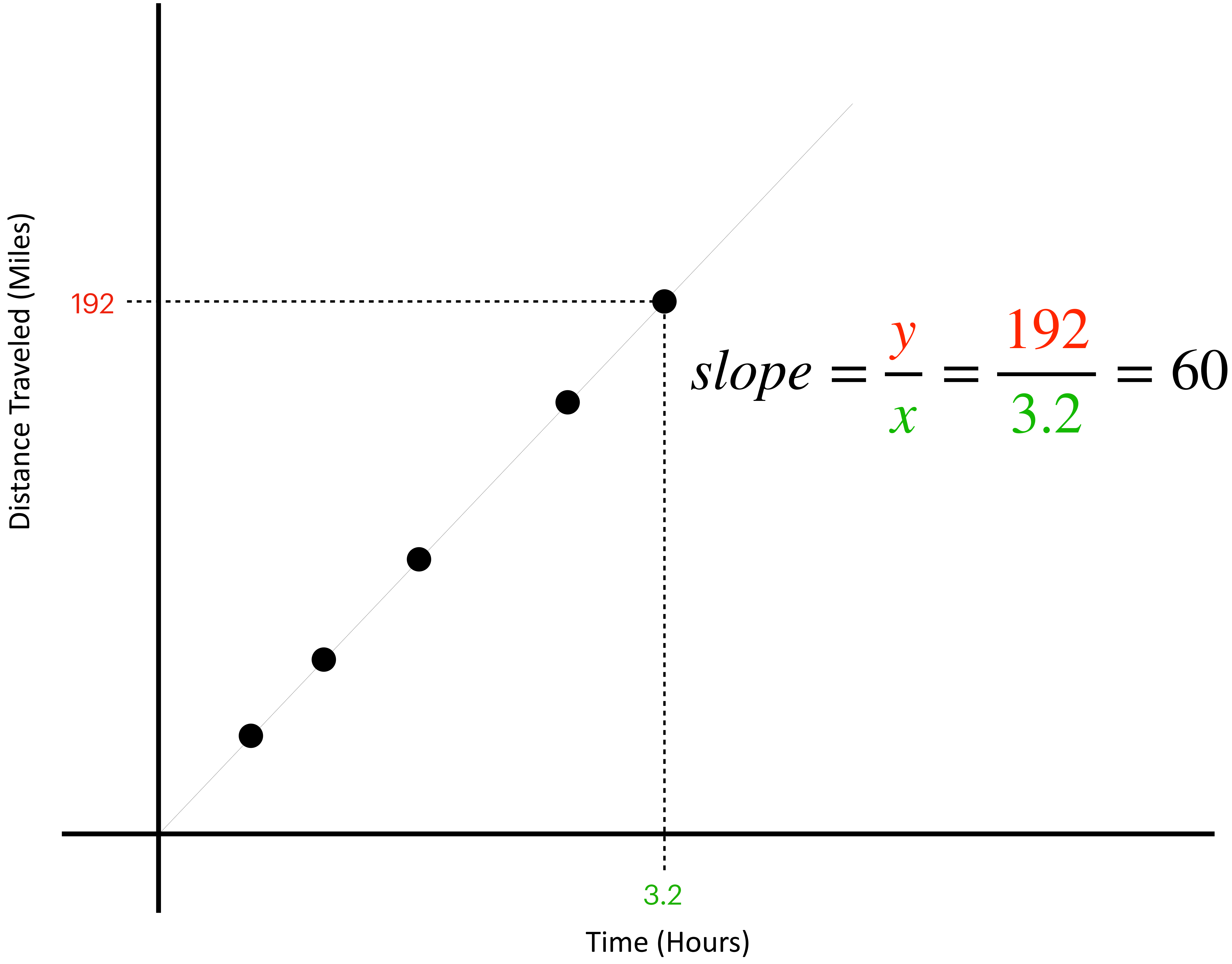


Simple Linear Regression

The slope of the line (i.e. speed of the car) is easy to determine:

Slope at $x = 3.2$, $y = 192$

Time (Hours)	Distance Traveled (Miles)	Speed (mph)
0.3	18	18 / 0.3 = 60
0.7	42	42 / 0.7 = 60
1.3	78	78 / 1.3 = 60
2.4	144	144 / 2.4 = 60
3.2	192	192 / 3.2 = 60



$$slope = \frac{y}{x} = \frac{192}{3.2} = 60$$

Simple Linear Regression

Once we know the slope of the line…

$$slope = 60$$

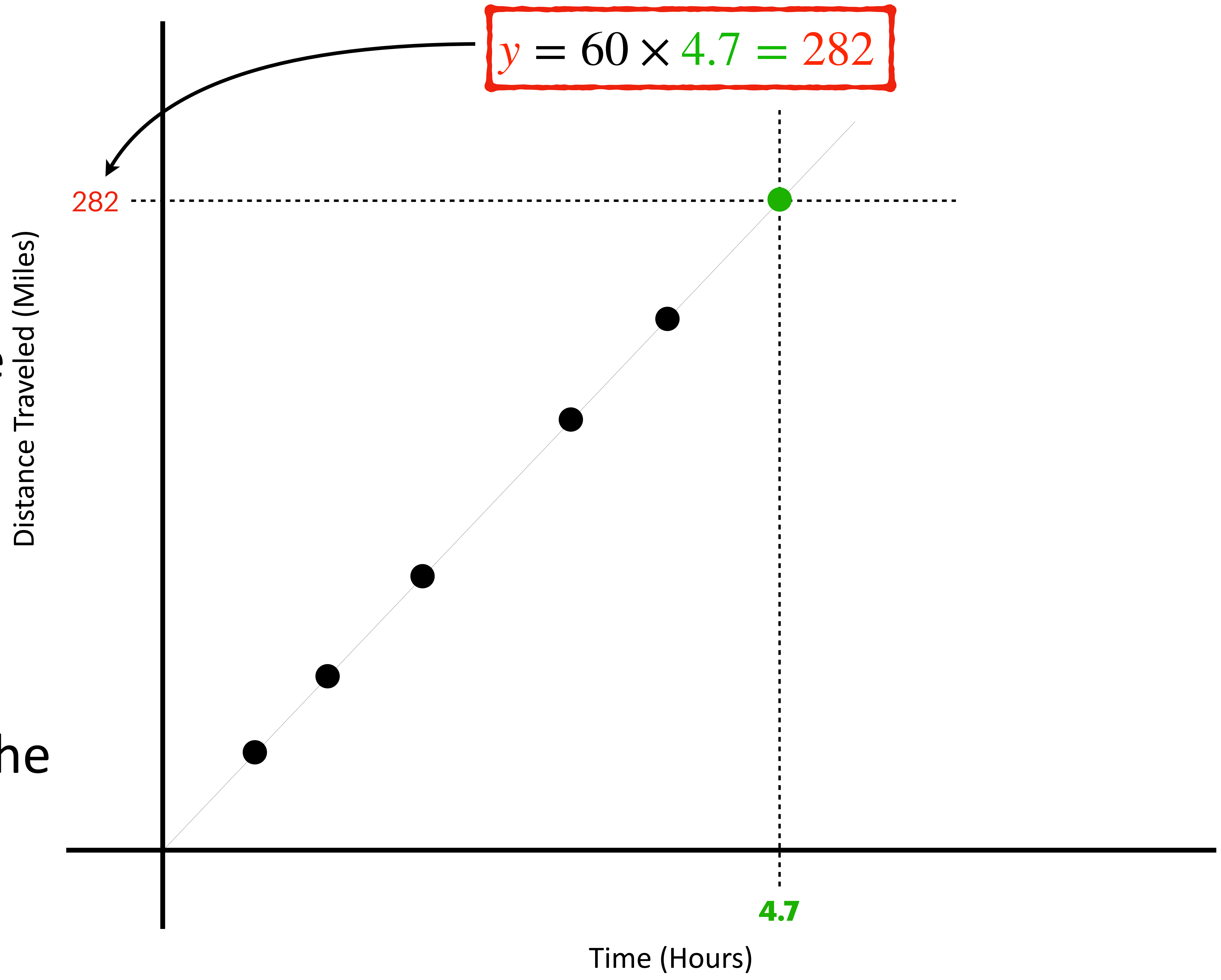
- Once we know the slope of the line, we can plot it using the formula for a line

$$y = \beta x$$

$$distance = speed \times time$$

- Once we have a line then we can find the distance traveled at any point in time

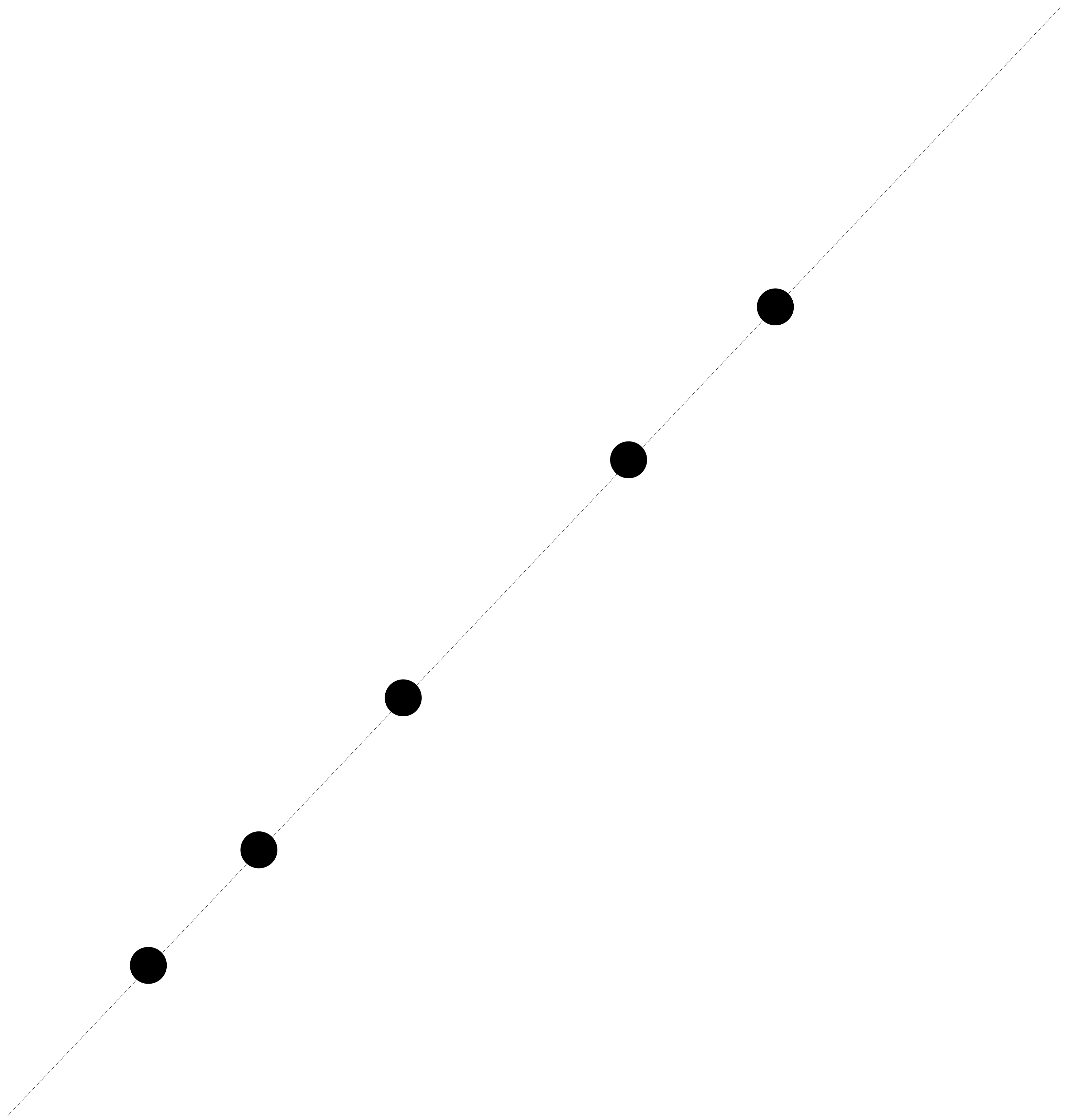
$$y = 60 \times 4.7 = 282$$



Simple Linear Regression

In General…

- Given a set of data points (x, y) …
- We plot a line that fits that data…
- Then we use the line to calculate the y values for any value of x



This was a simple (contrived?) example…

- All the data points lined up perfectly
- The line fit the data perfectly
- We could have simply used the formula for the distance

$$distance = speed \times time$$

Lets take a more realistic example…

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102

Question: Can we **predict** the weight of a person that is 71 inches tall?

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102

Question: Can we **predict** the weight of a person that is 71 inches tall?

Weight (lbs)

Lets begin by plotting the data

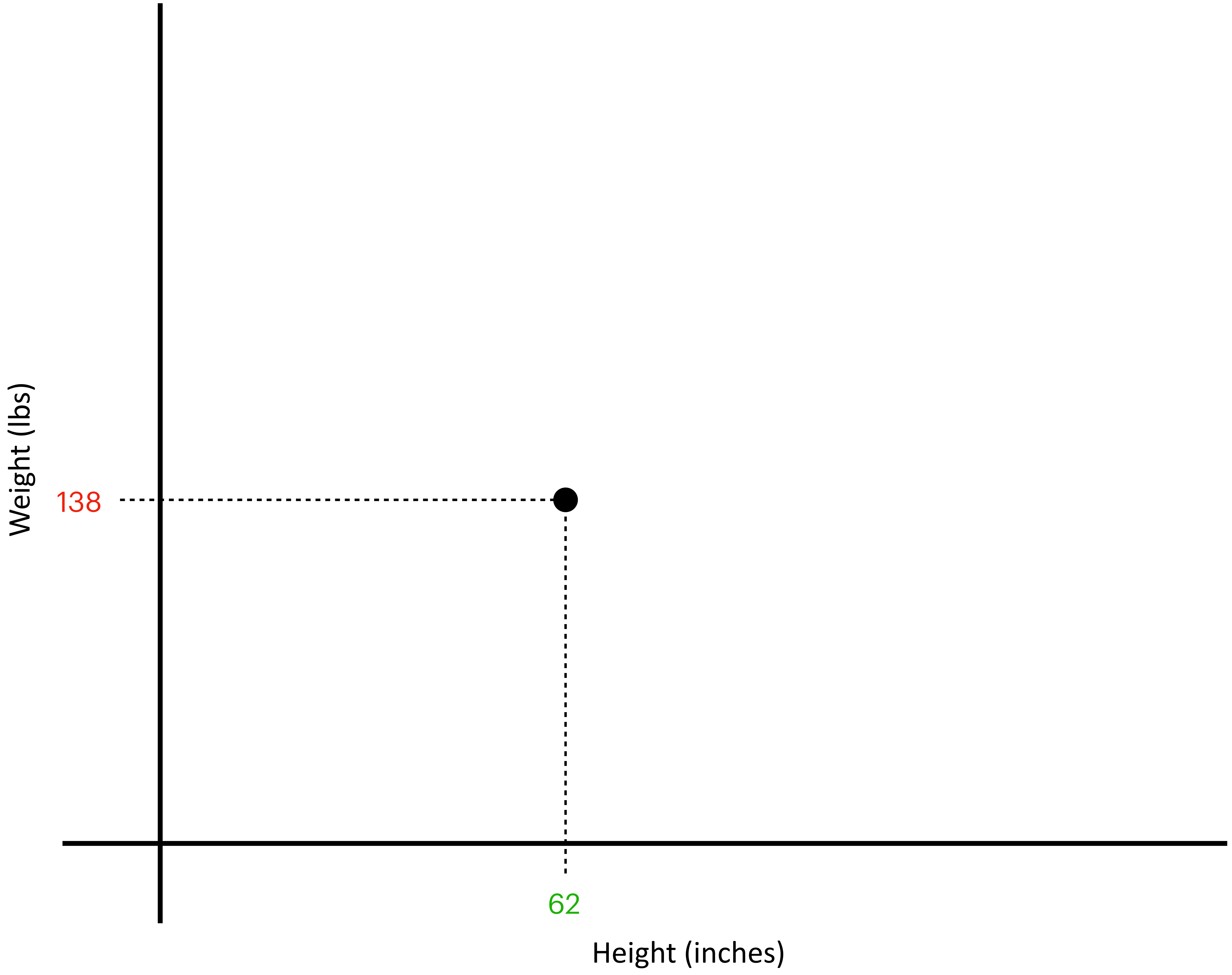
Y Axis = Weight (lbs)
X Axis = Height (inches)

Height (inches)

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	**138**
55	178
44	123
75	200
65	229
50	102



Question: Can we **predict** the weight of a person that is 71 inches tall?

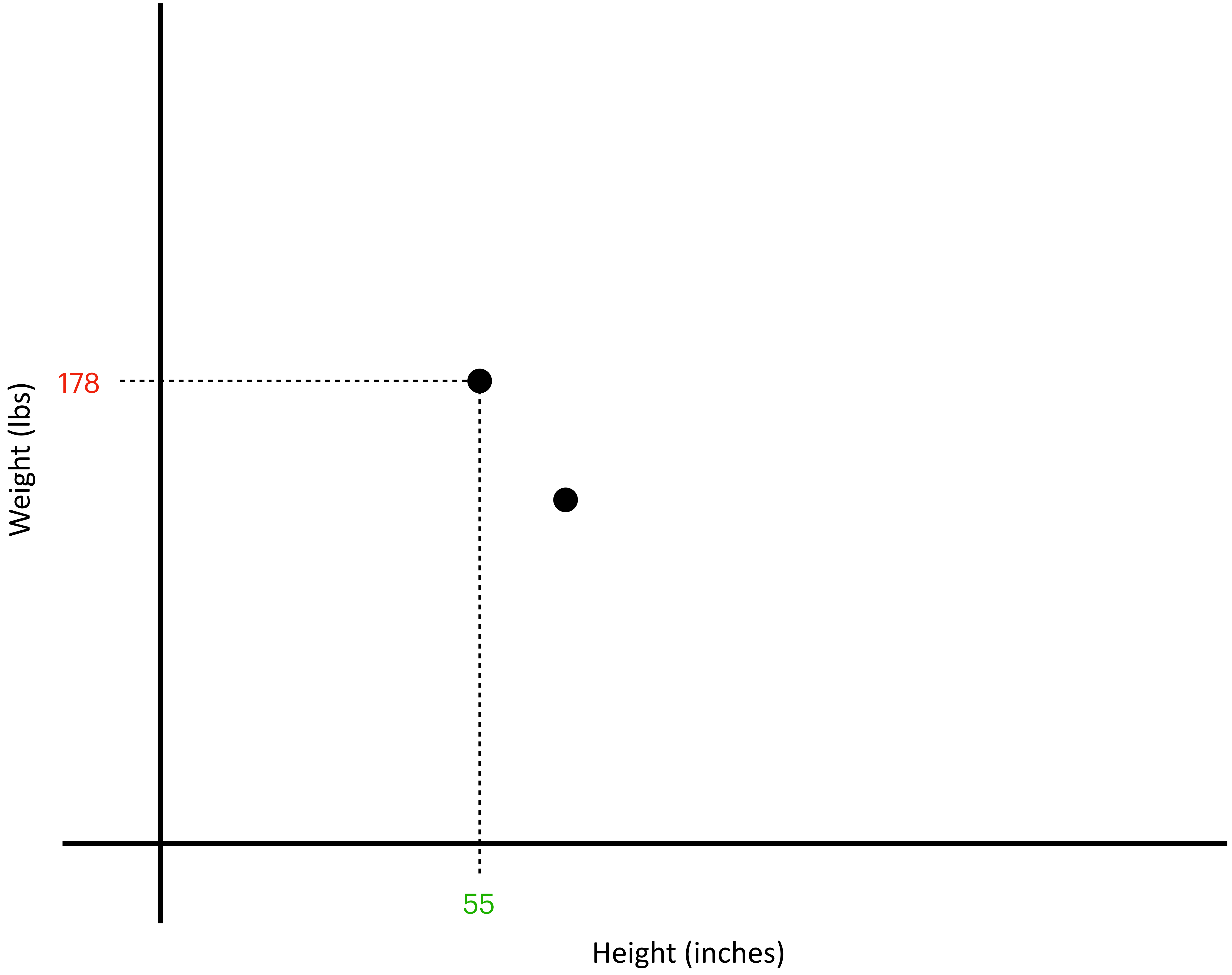
Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	**138**
55	178
44	123
75	200
65	229
50	102

Question: Can we **predict** the weight of a person that is 71 inches tall?

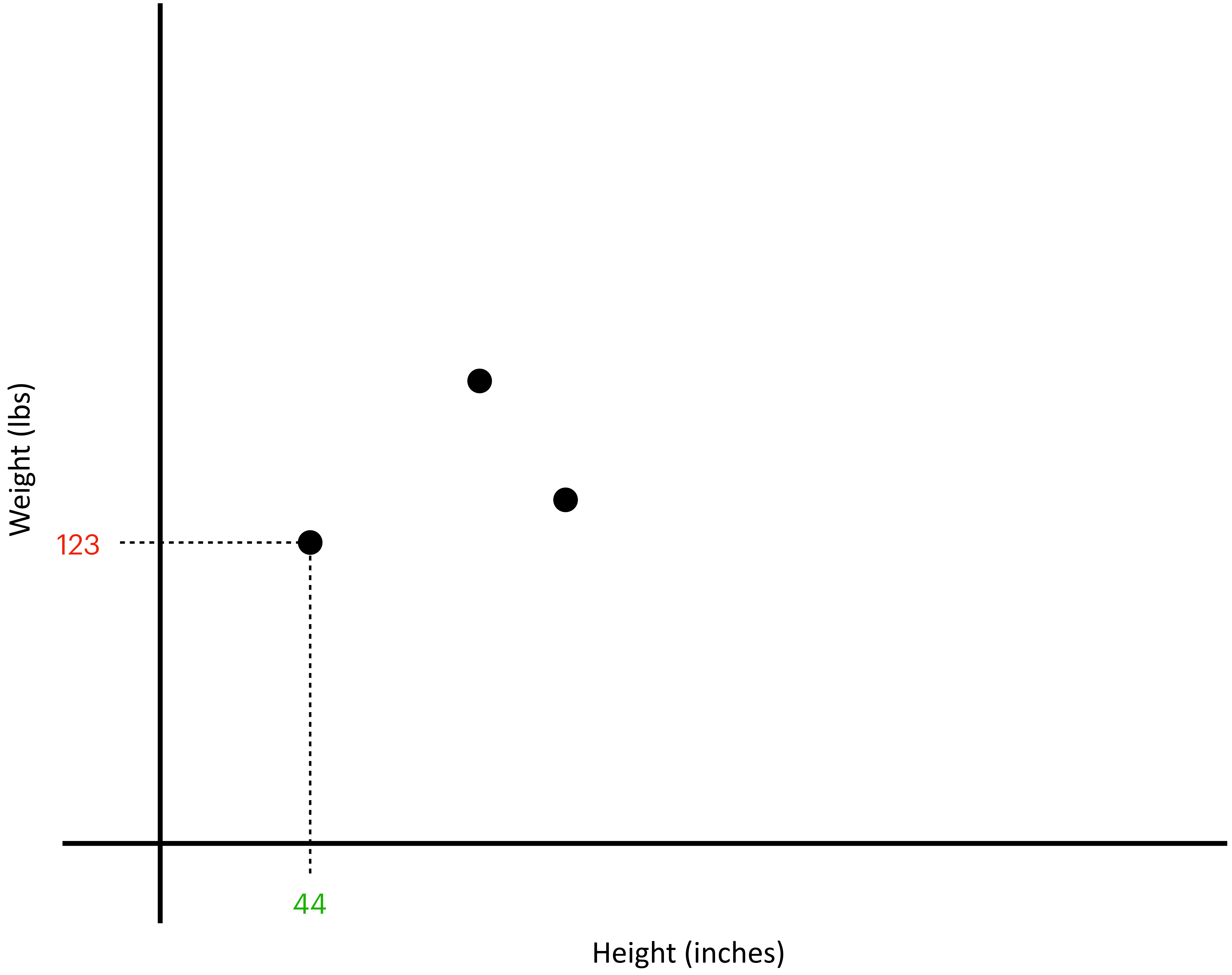


Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: Can we **predict** the weight of a person that is 71 inches tall?

Data is synthetic and not plotted to scale

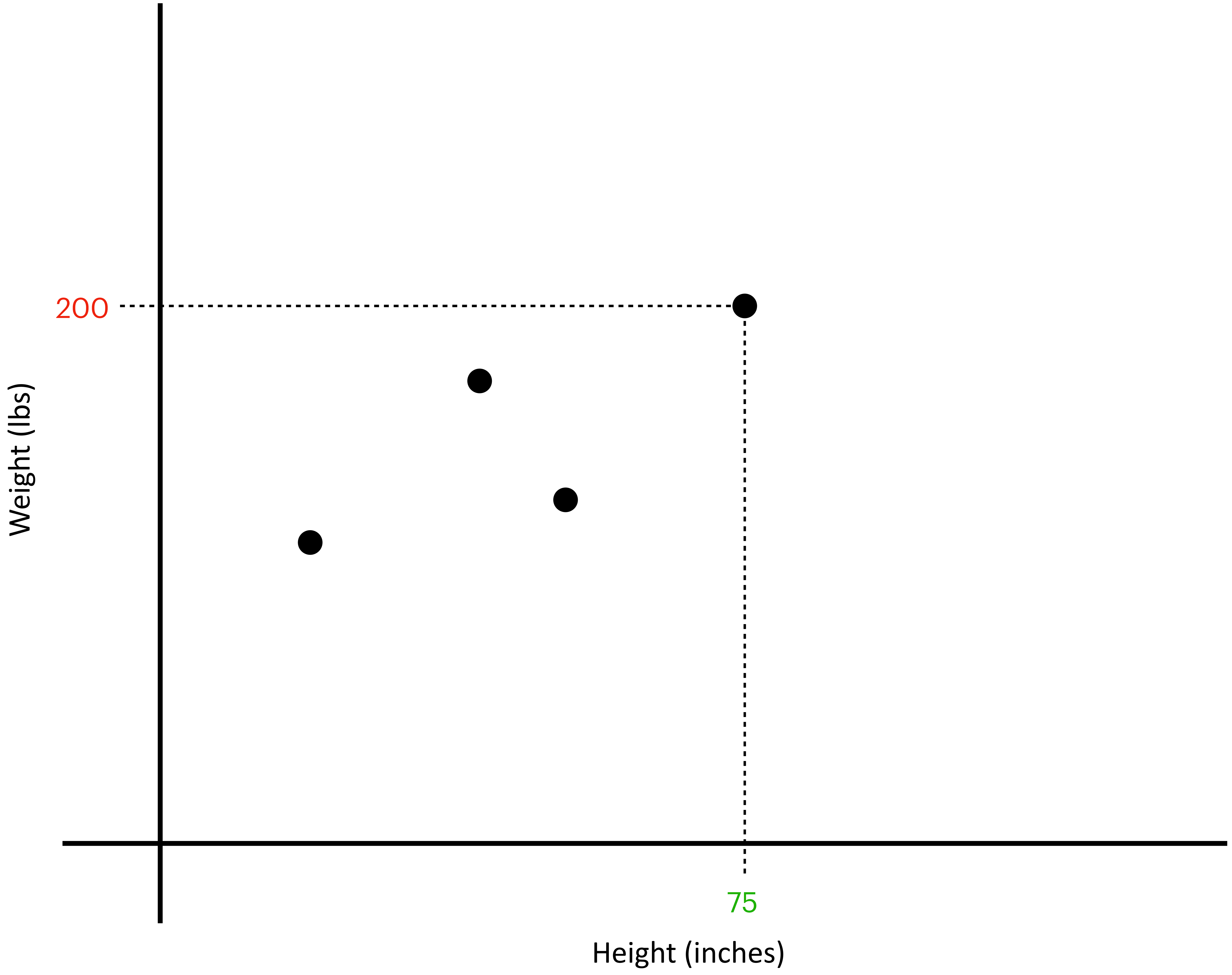
Simple Linear Regression

A realistic example...

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	**200**
65	229
50	102

Question: Can we **predict** the weight of a person that is 71 inches tall?



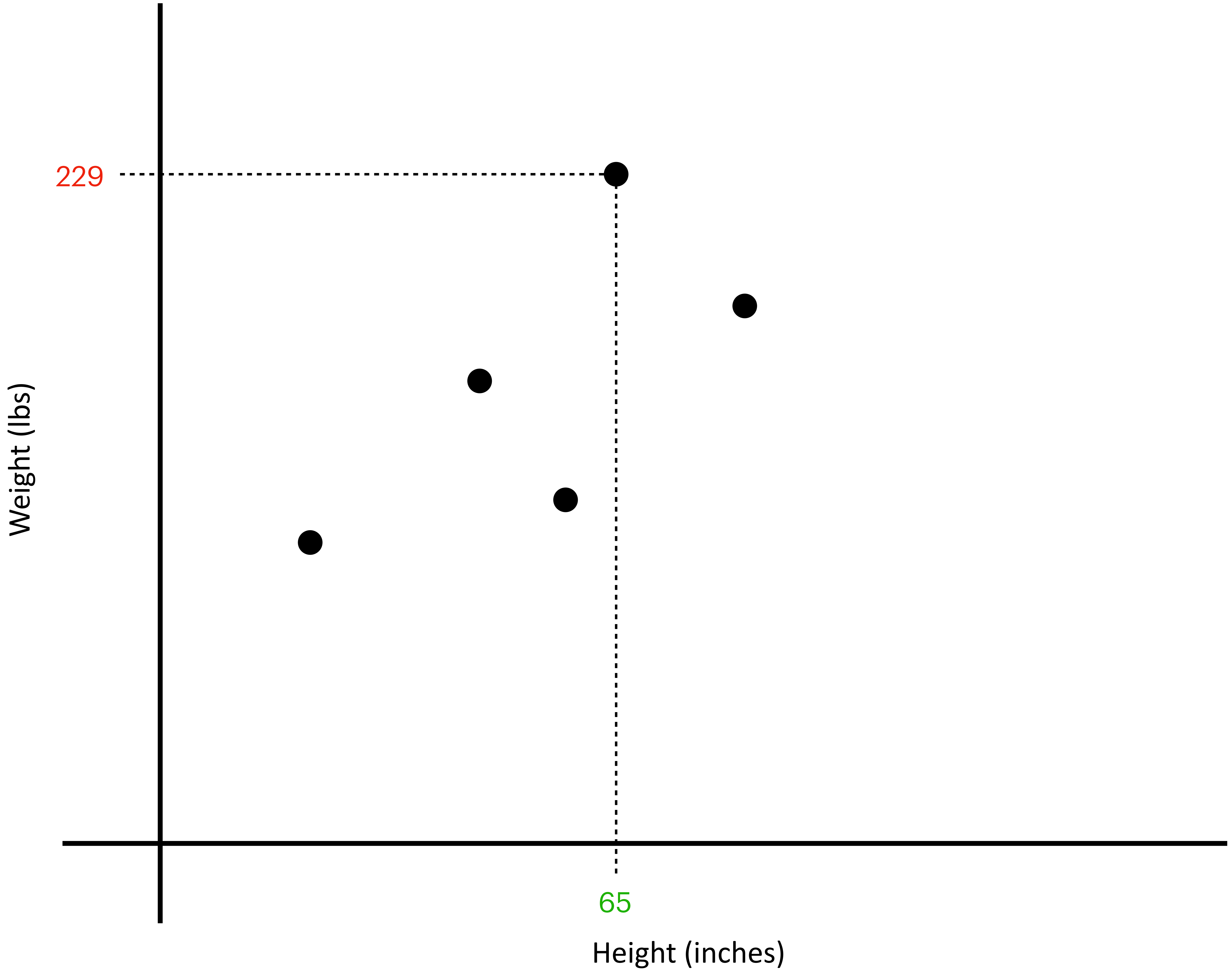
Data is synthetic and not plotted to scale

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	**229**
50	102



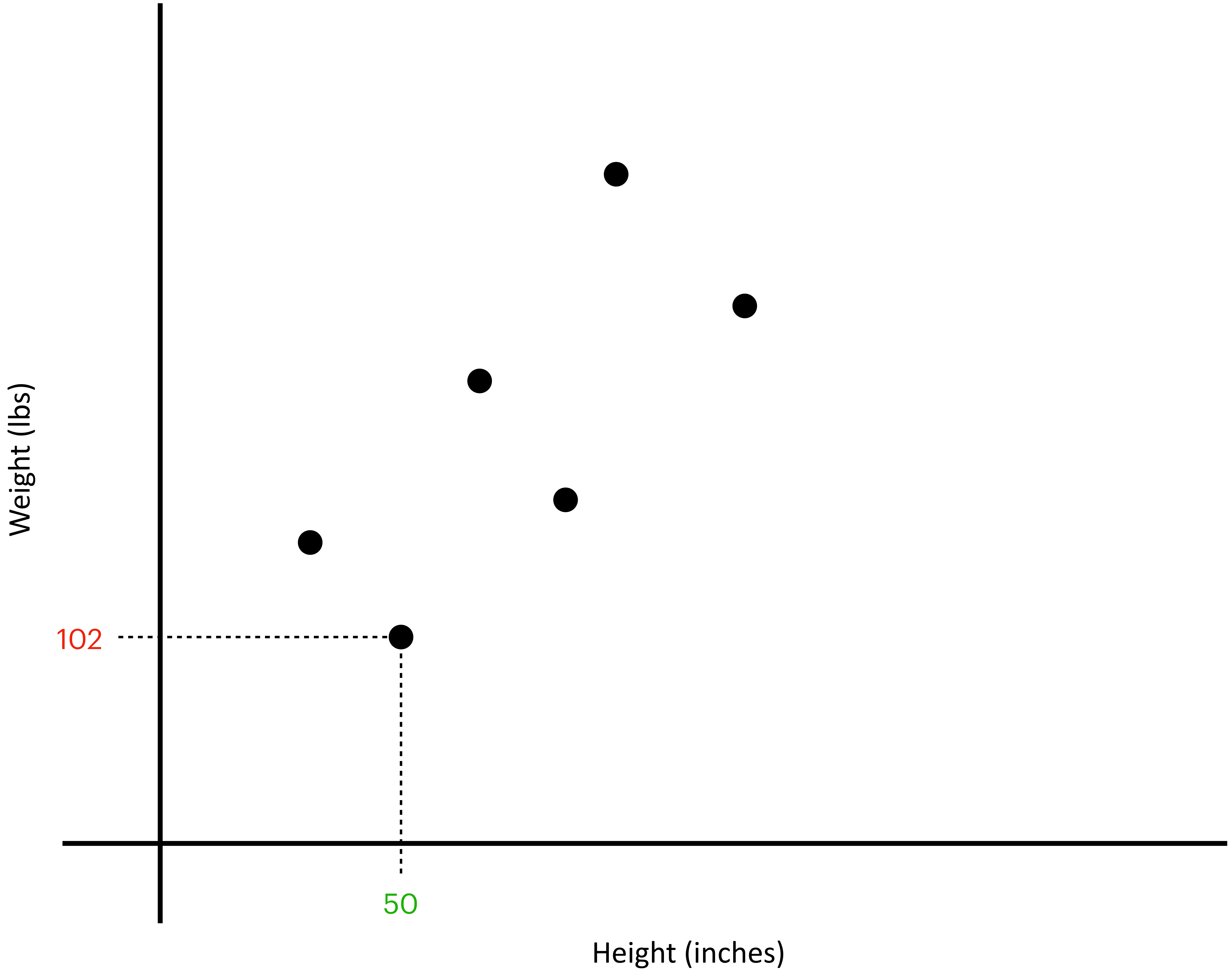
Question: Can we **predict** the weight of a person that is 71 inches tall?

Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



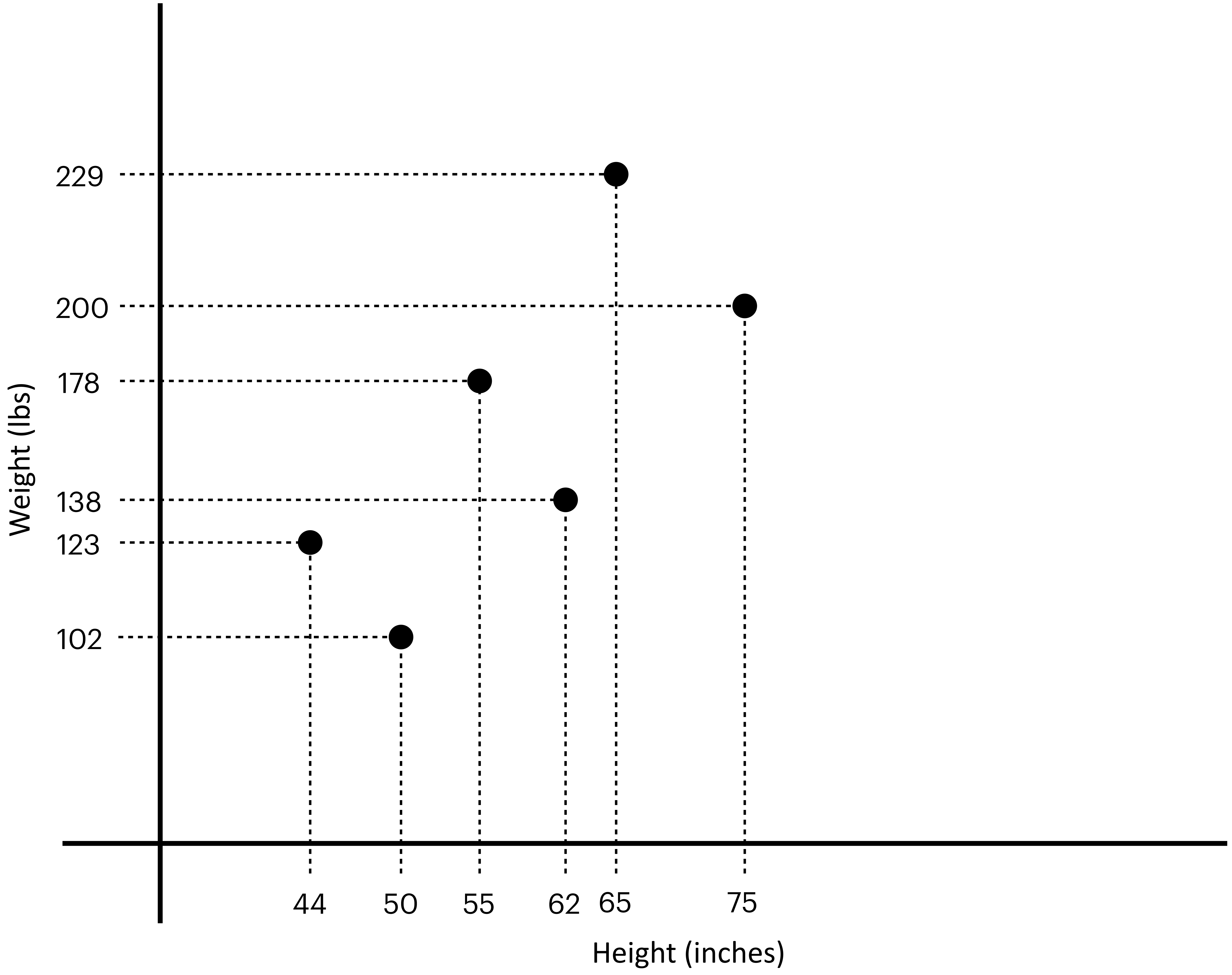
Question: Can we **predict** the weight of a person that is 71 inches tall?

Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: Can we **predict** the weight of a person that is 71 inches tall?

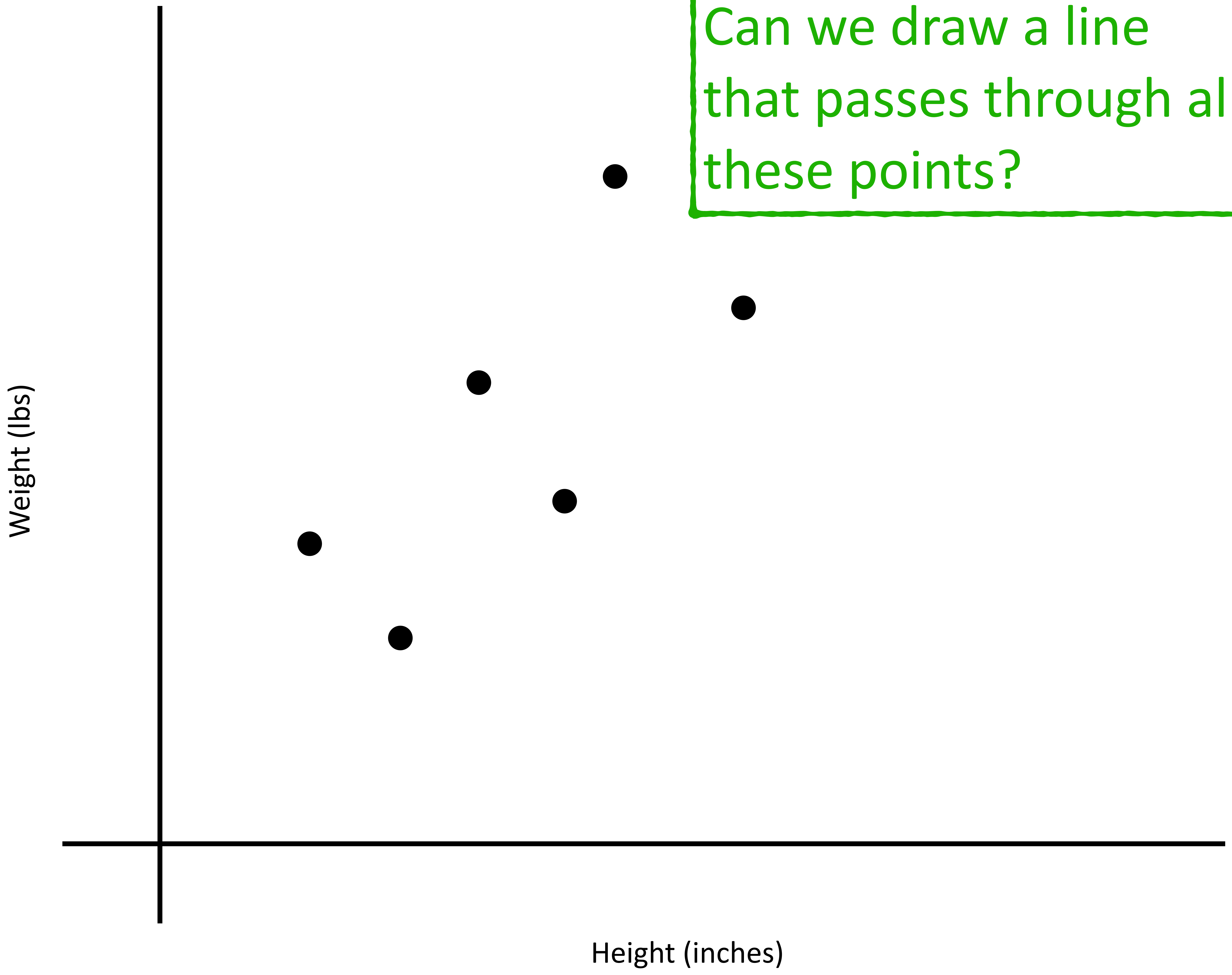
Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102

Question: Can we **predict** the weight of a person that is 71 inches tall?



Can we draw a line that passes through all these points?

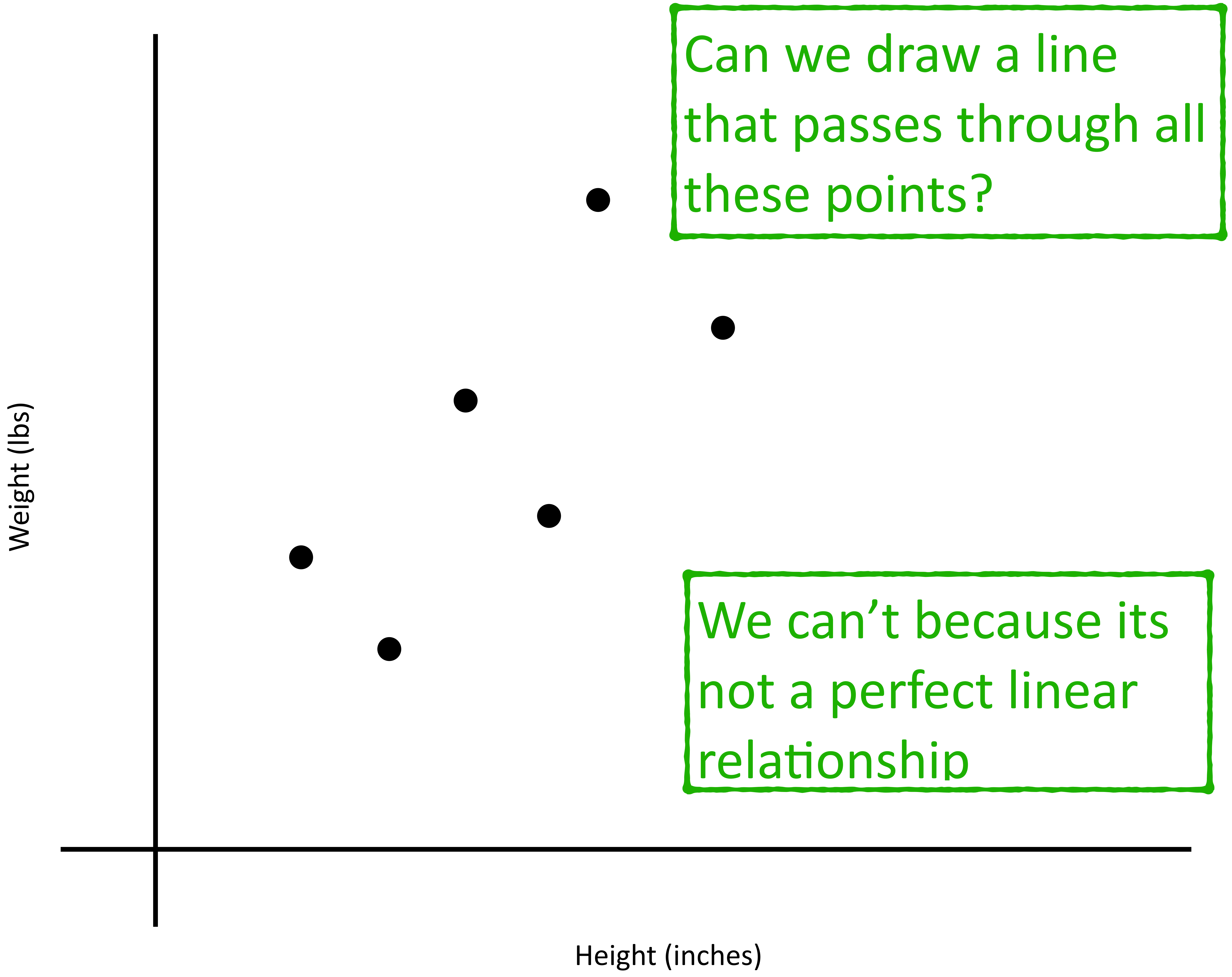
Data is synthetic and not plotted to scale

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: Can we **predict** the weight of a person that is 71 inches tall?

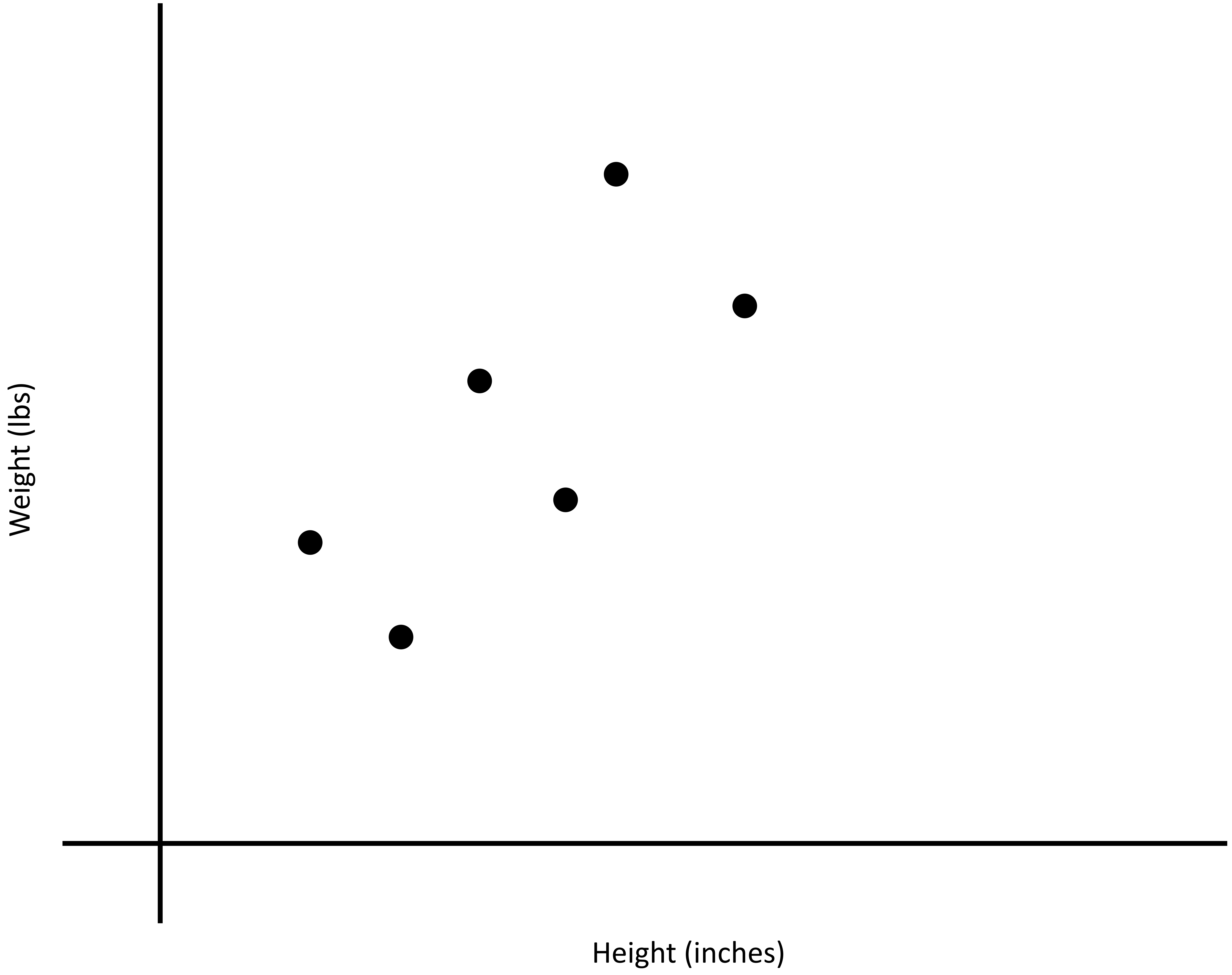
Data is synthetic and not plotted to scale

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: So what's the line that **best** fits the data?

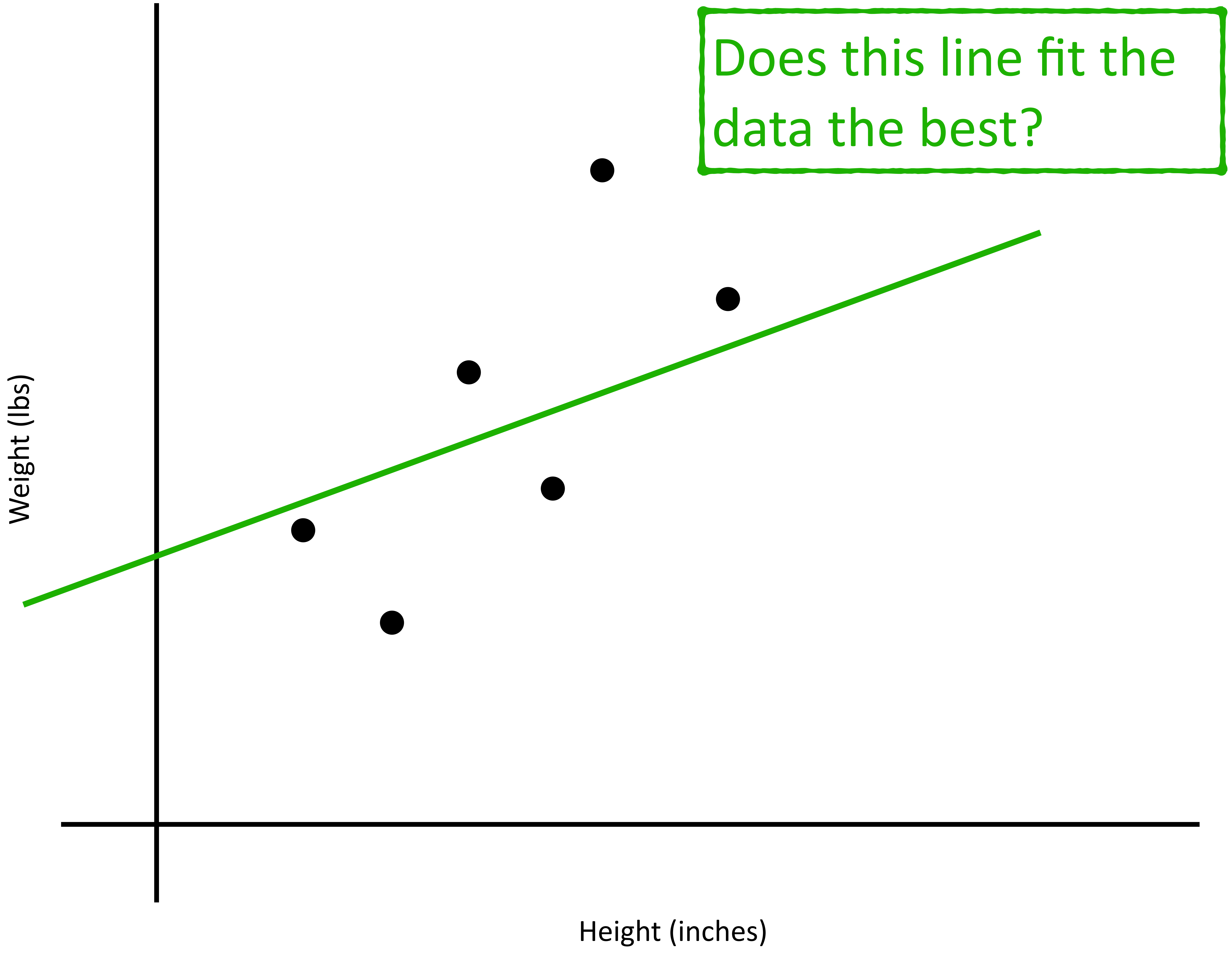
Data is synthetic and not plotted to scale

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: So what's the line that **best** fits the data?

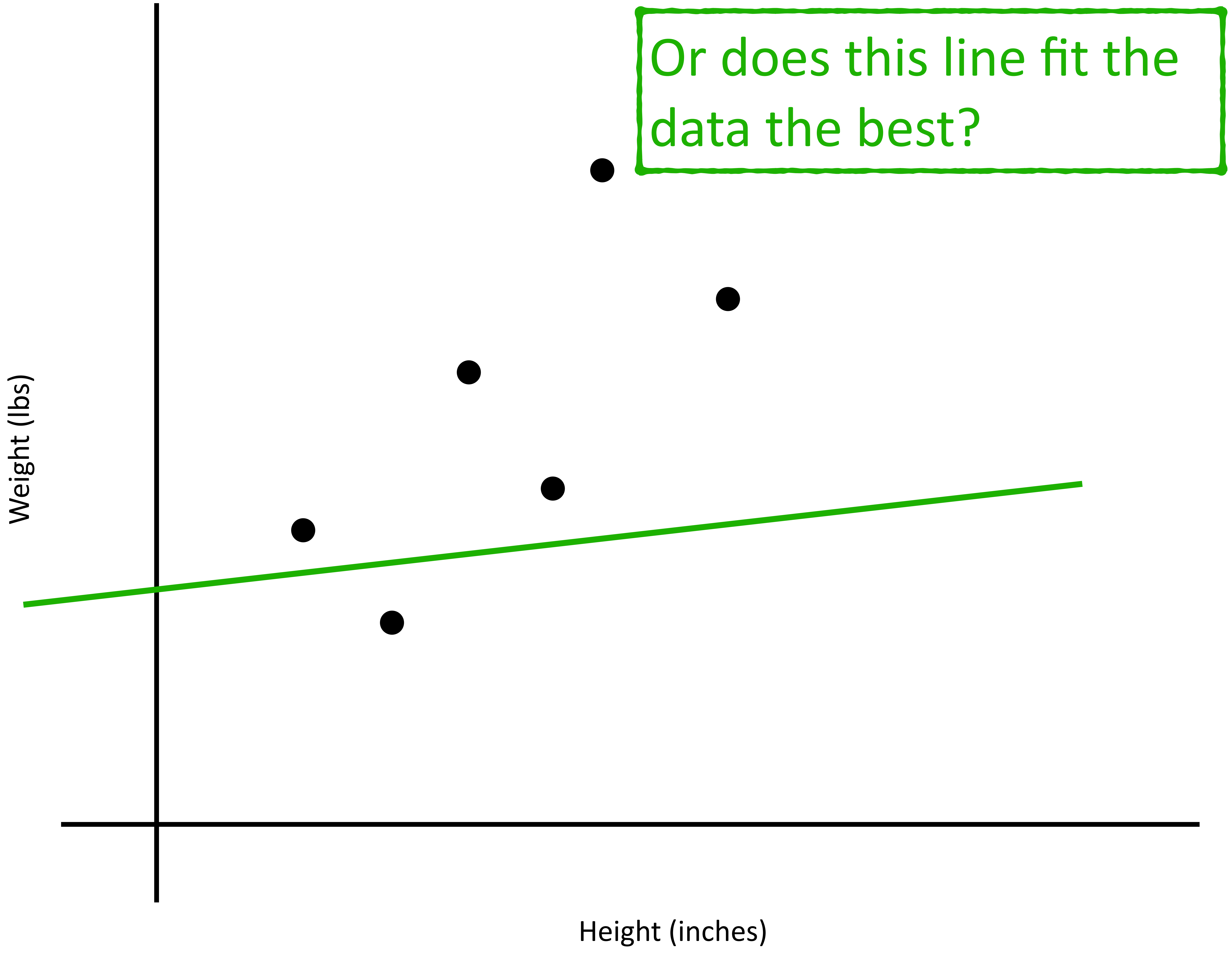
Data is synthetic and not plotted to scale

Simple Linear Regression

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: So what's the line that **best** fits the data?

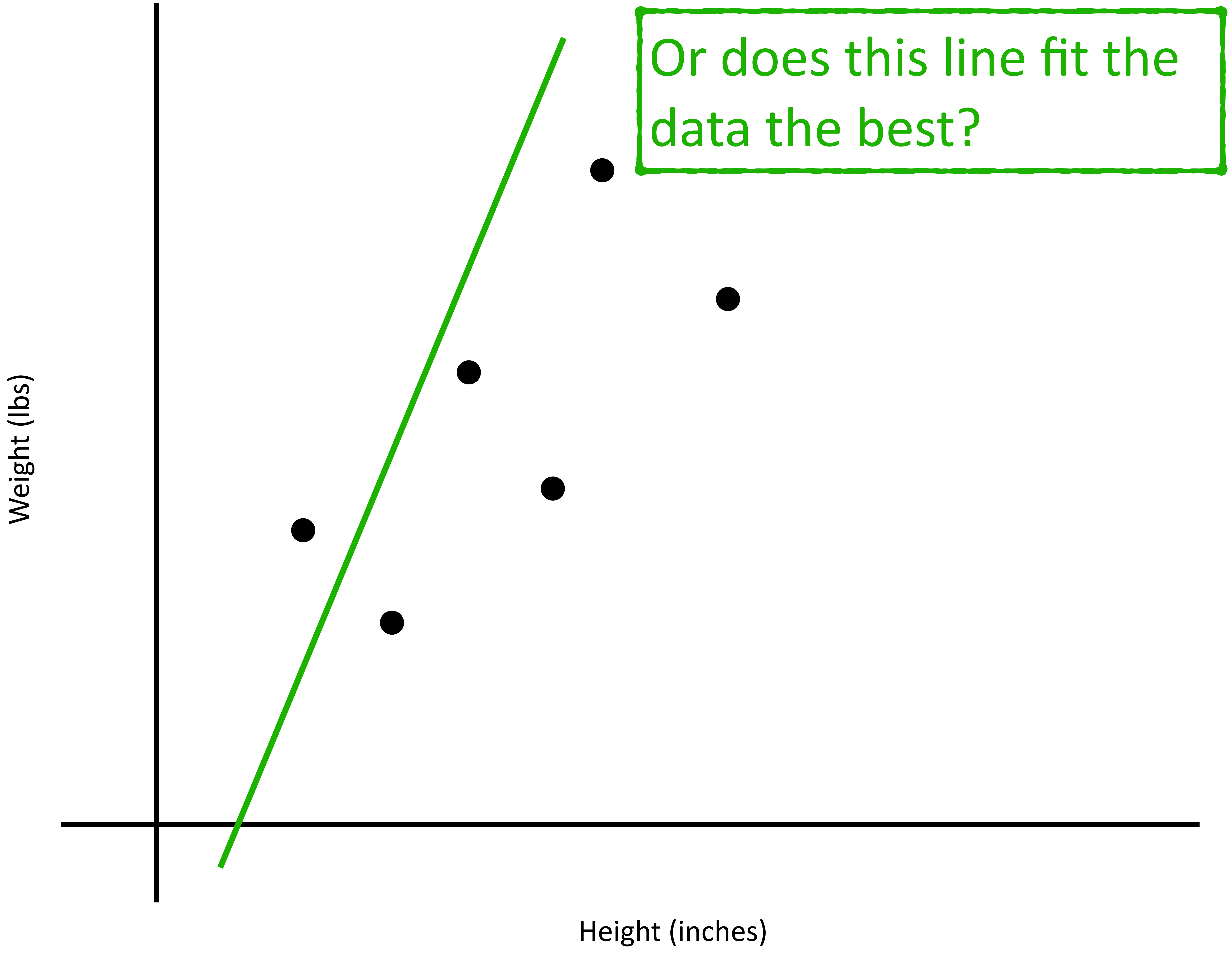
Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102

Question: So what's the line that **best** fits the data?

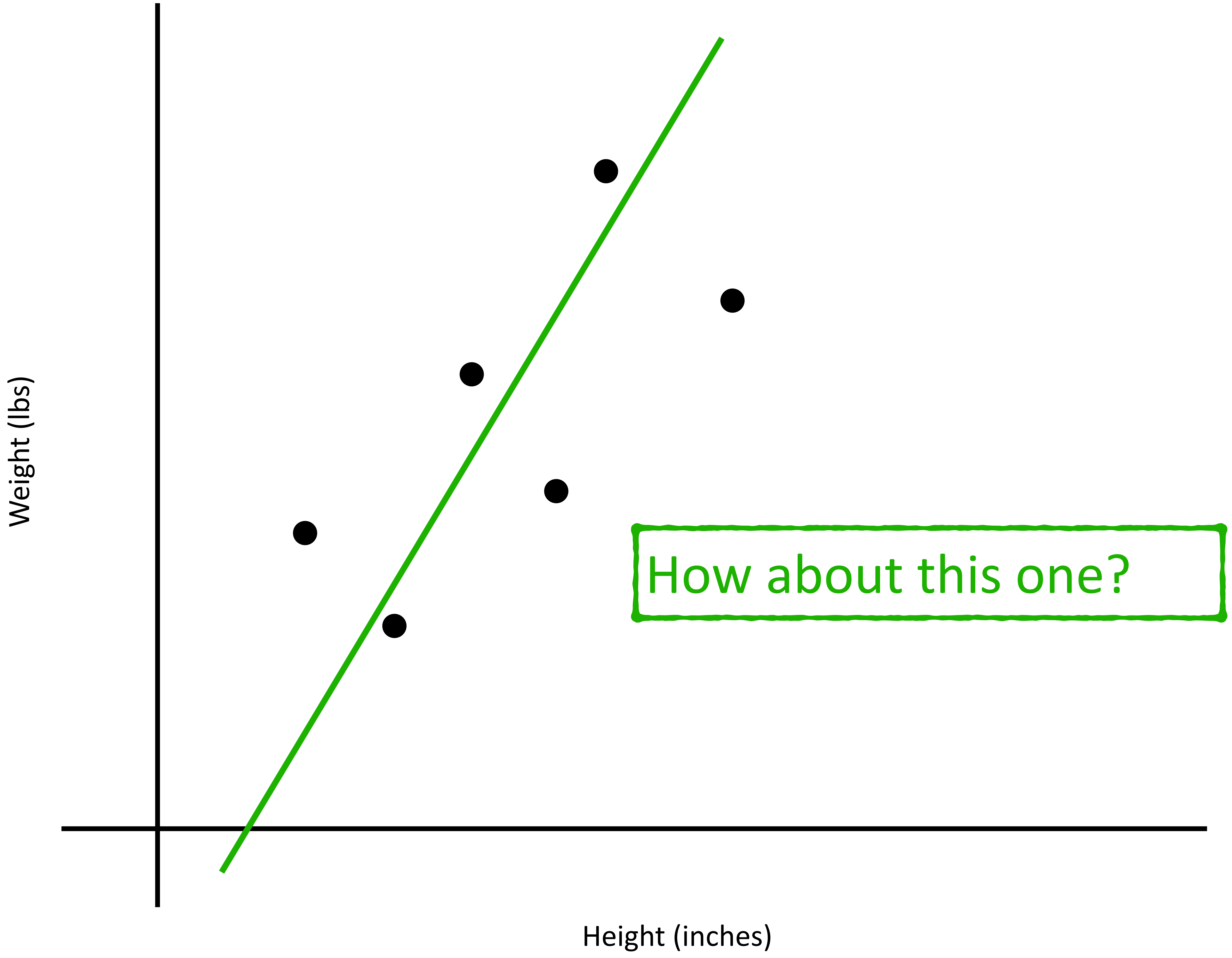


Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



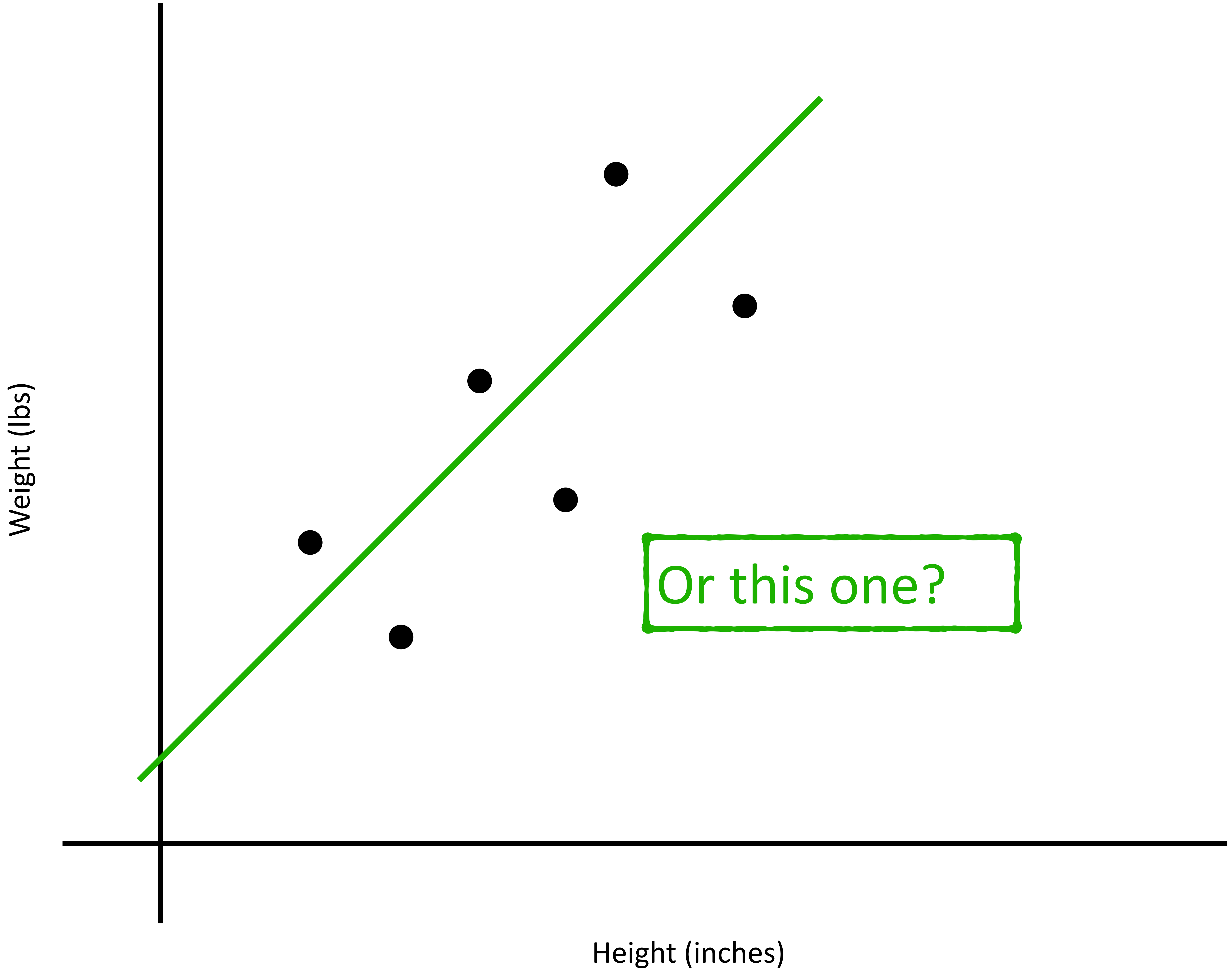
Question: So what's the line that **best** fits the data?

Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



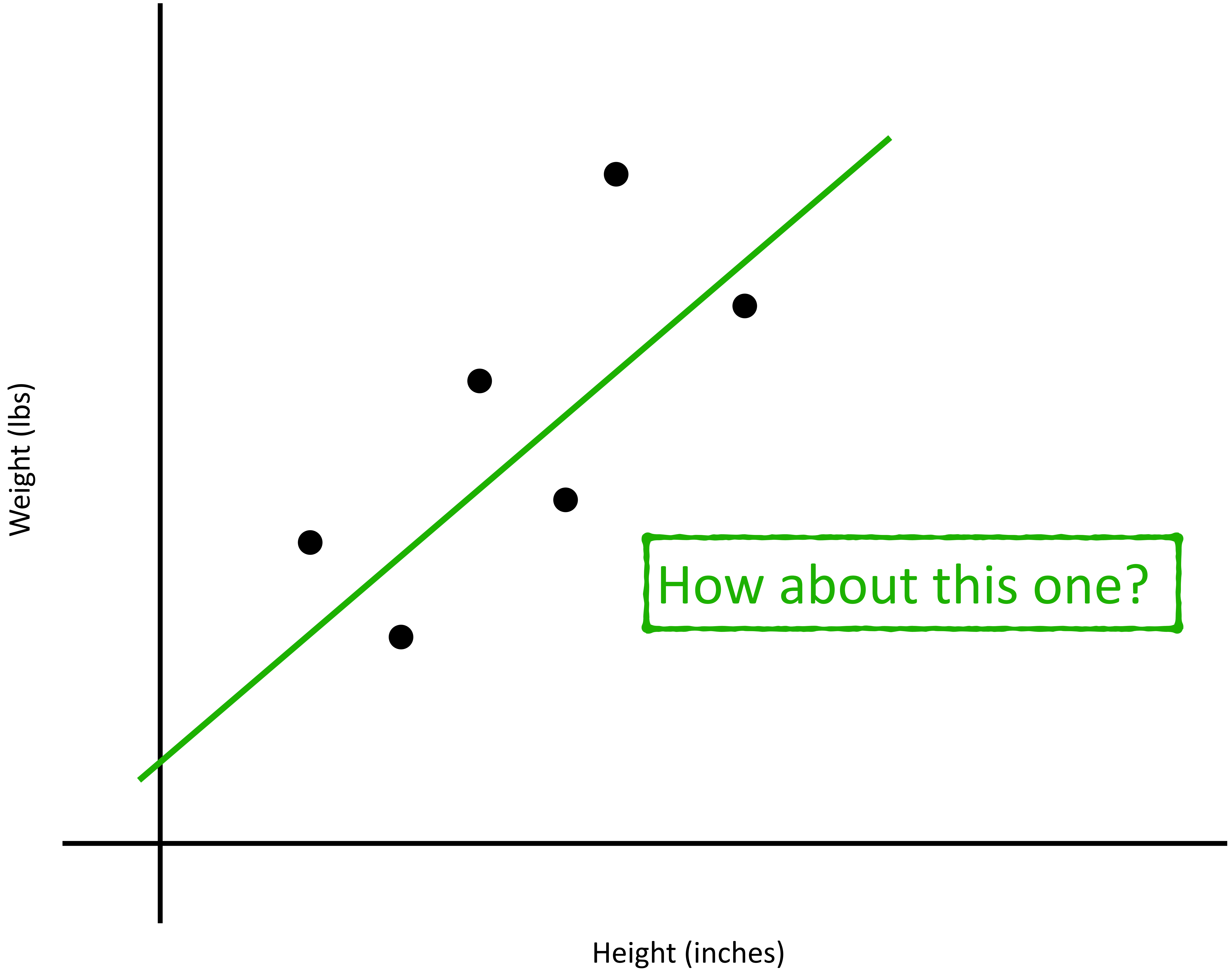
Question: So what's the line that **best** fits the data?

Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
62	138
55	178
44	123
75	200
65	229
50	102



Question: So what's the line that **best** fits the data?

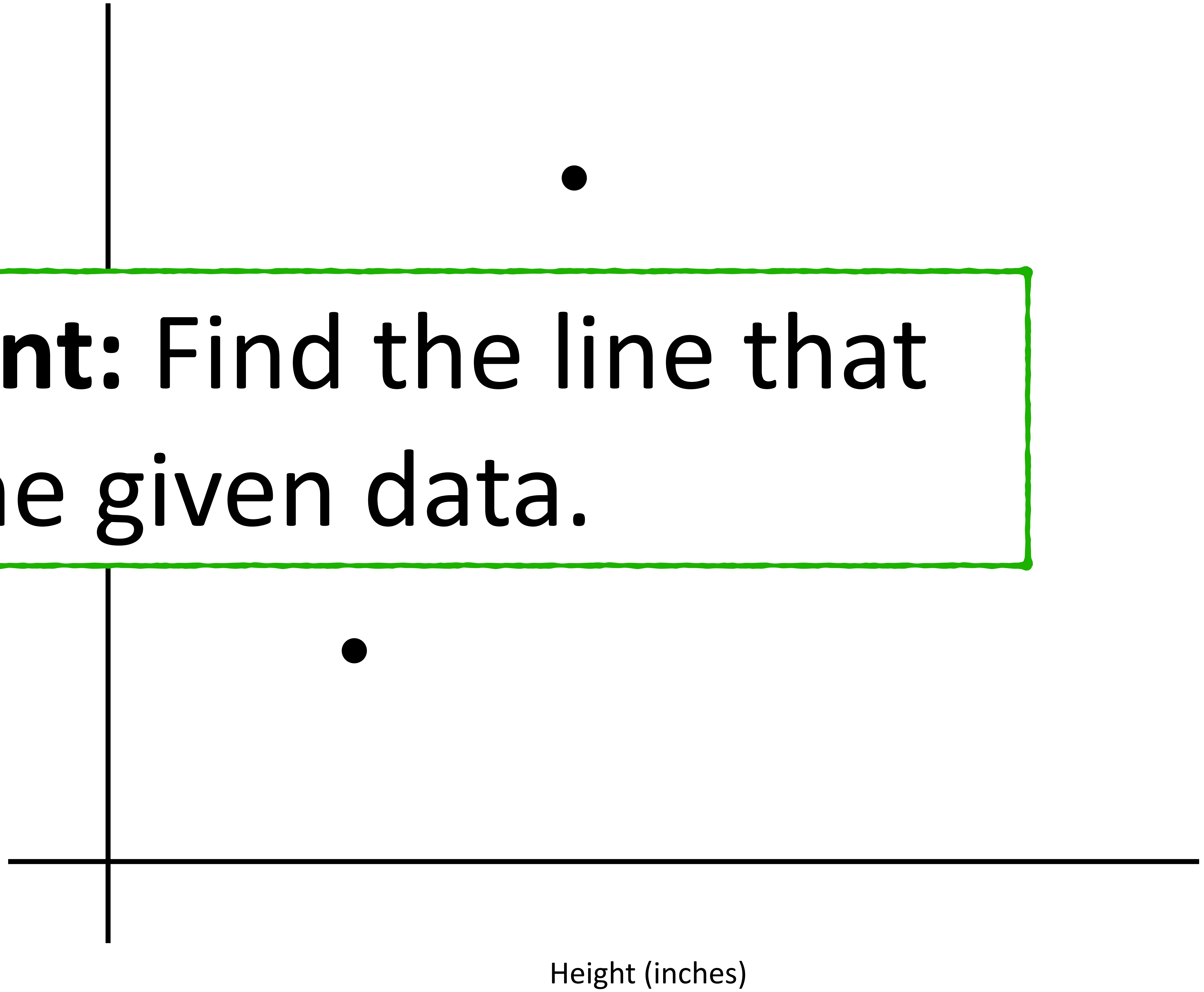
Data is synthetic and not plotted to scale

A realistic example…

We observe the heights and weights of 6 people

Height (inches)	Weight (lbs)
50	102

Problem Statement: Find the line that **best** fits the given data.

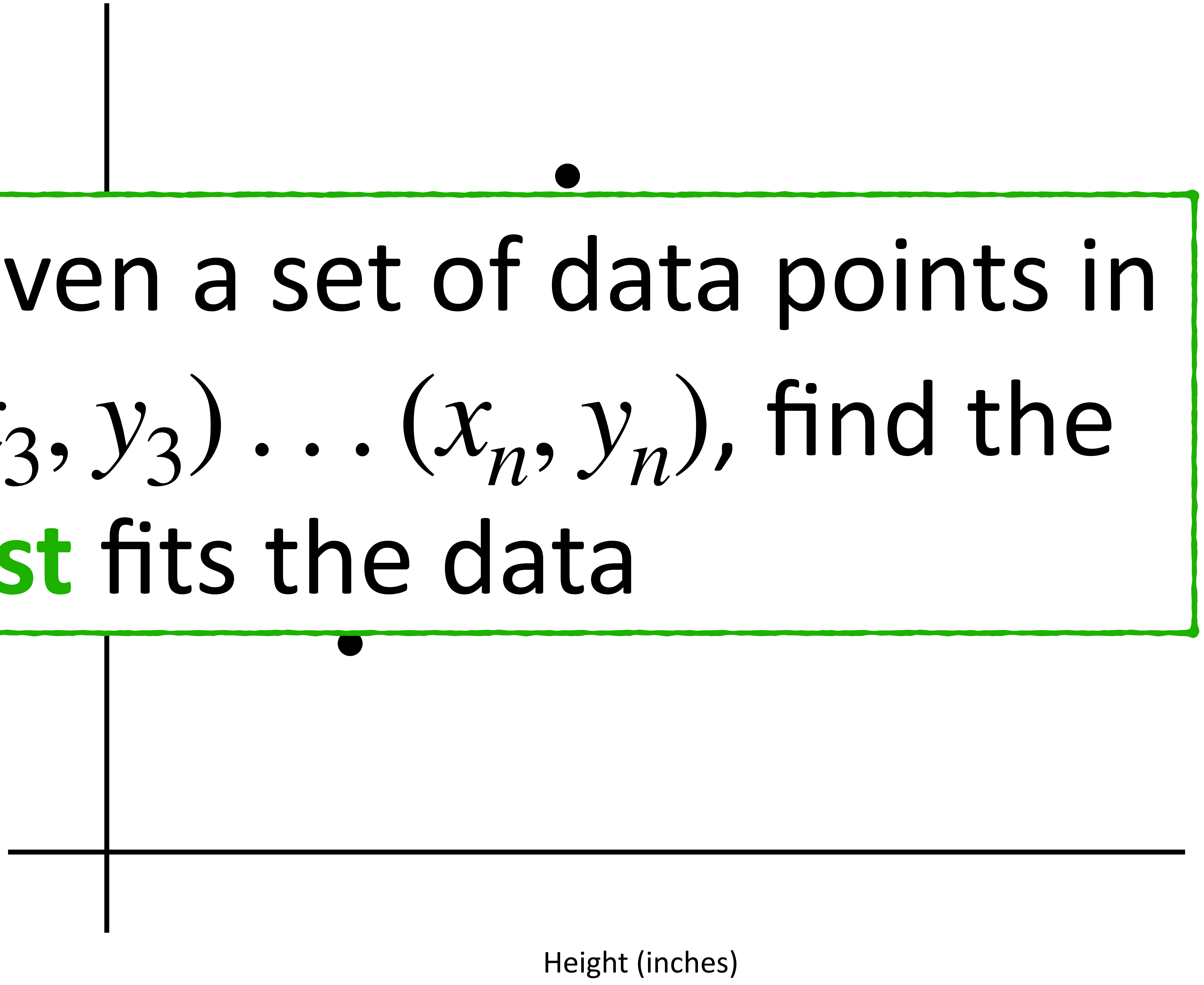


Data is synthetic and not plotted to scale

A realistic example...

We observe the heights and weights of 6 people

Problem Statement: Given a set of data points in $\mathbf{R}^2$, $(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots (x_n, y_n)$, find the line that **best** fits the data



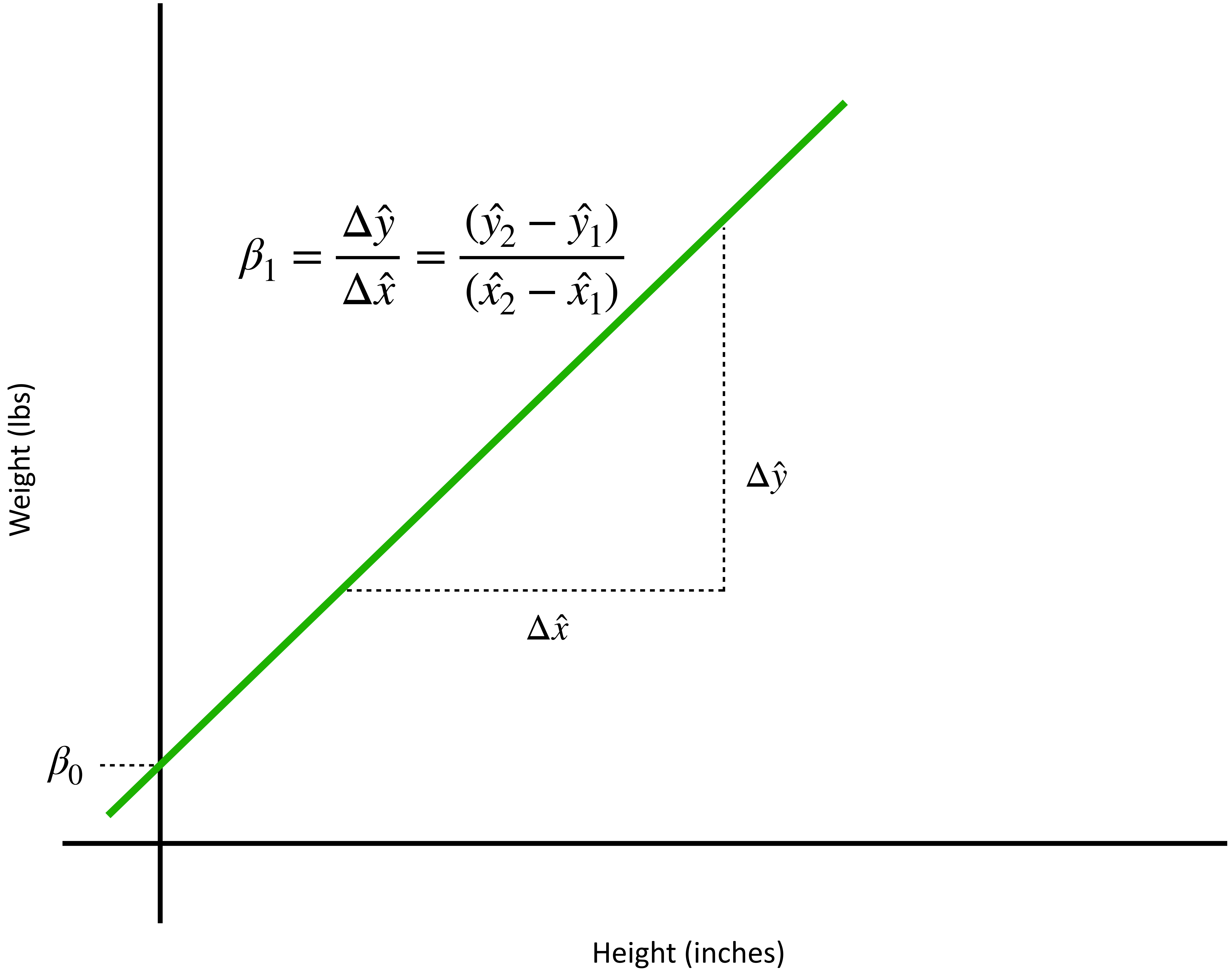
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

β_0 Is the Y intercept

β_1 Is the slope of the line



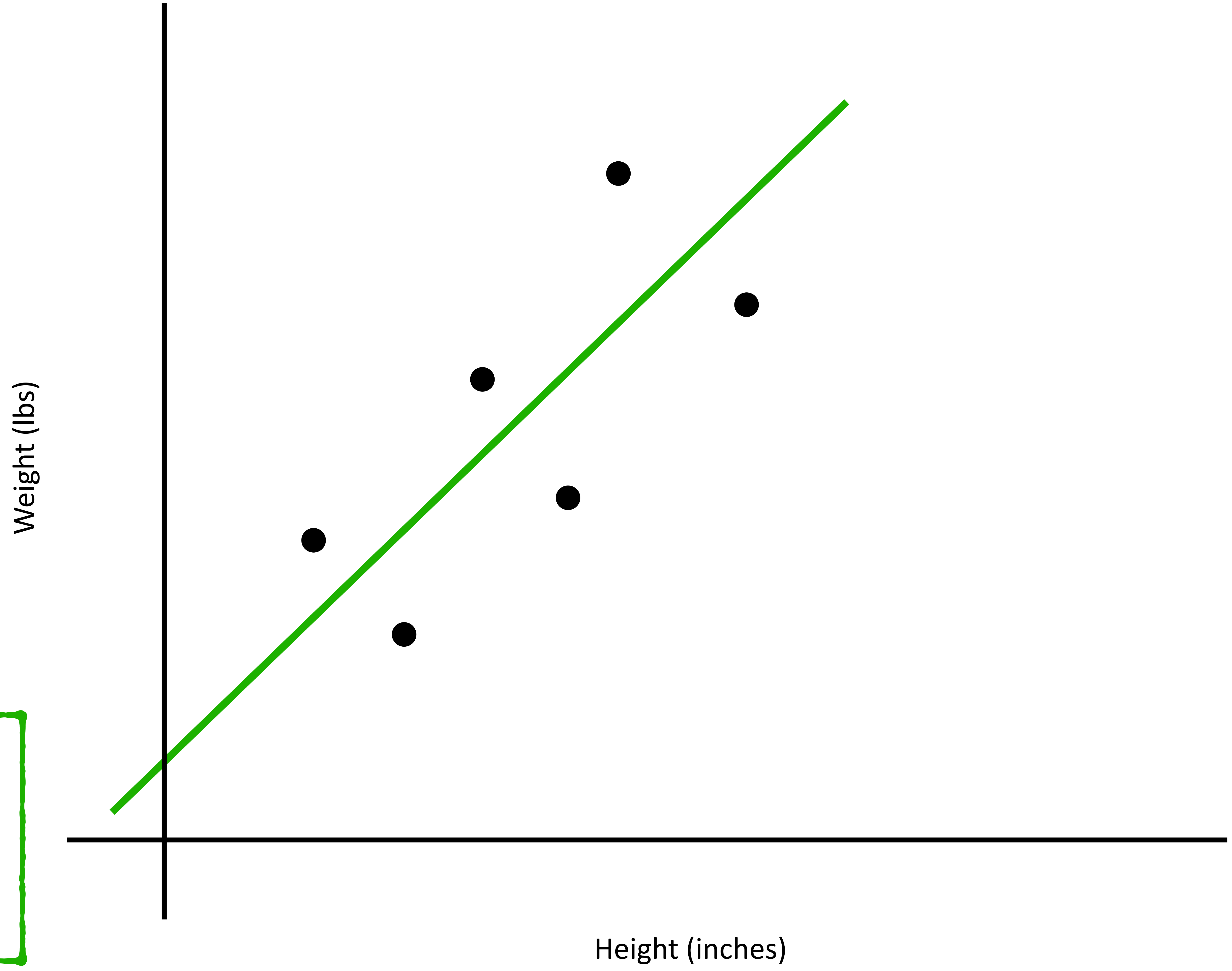
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

β_0 Is the Y intercept

β_1 Is the slope of the line

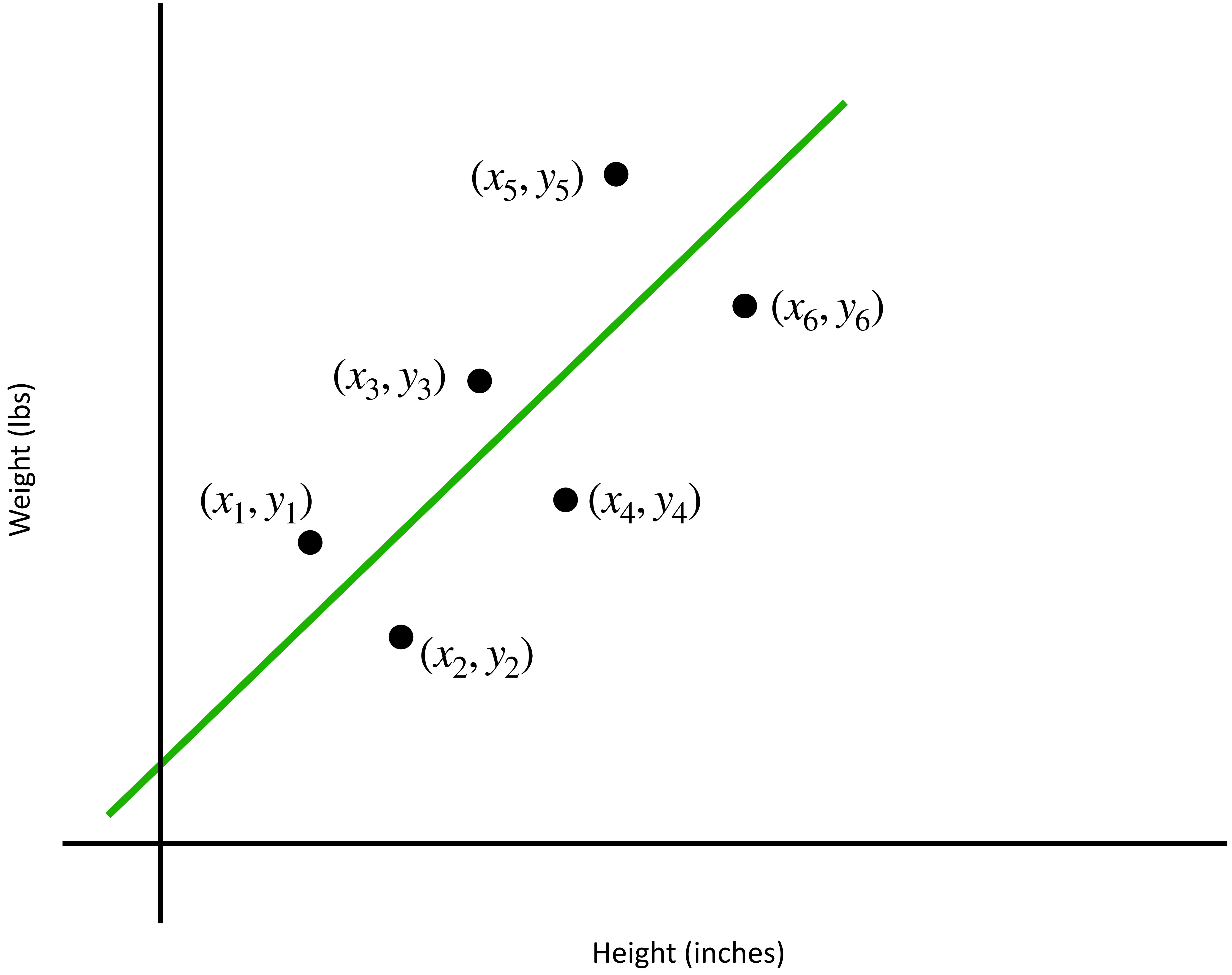


Problem Statement: Find the values of β_0 and β_1 for the line that best fits the given data.

Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$



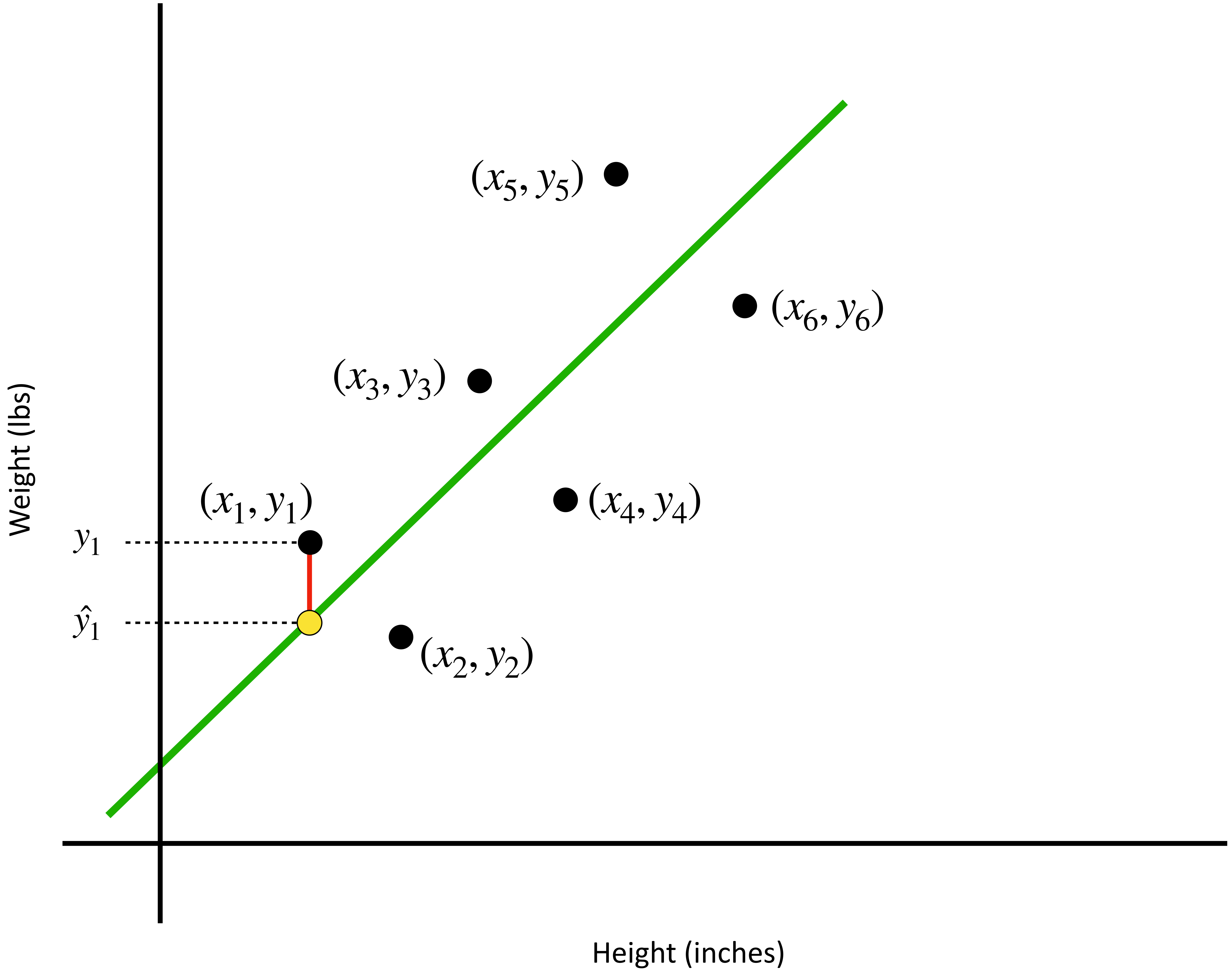
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1)$$



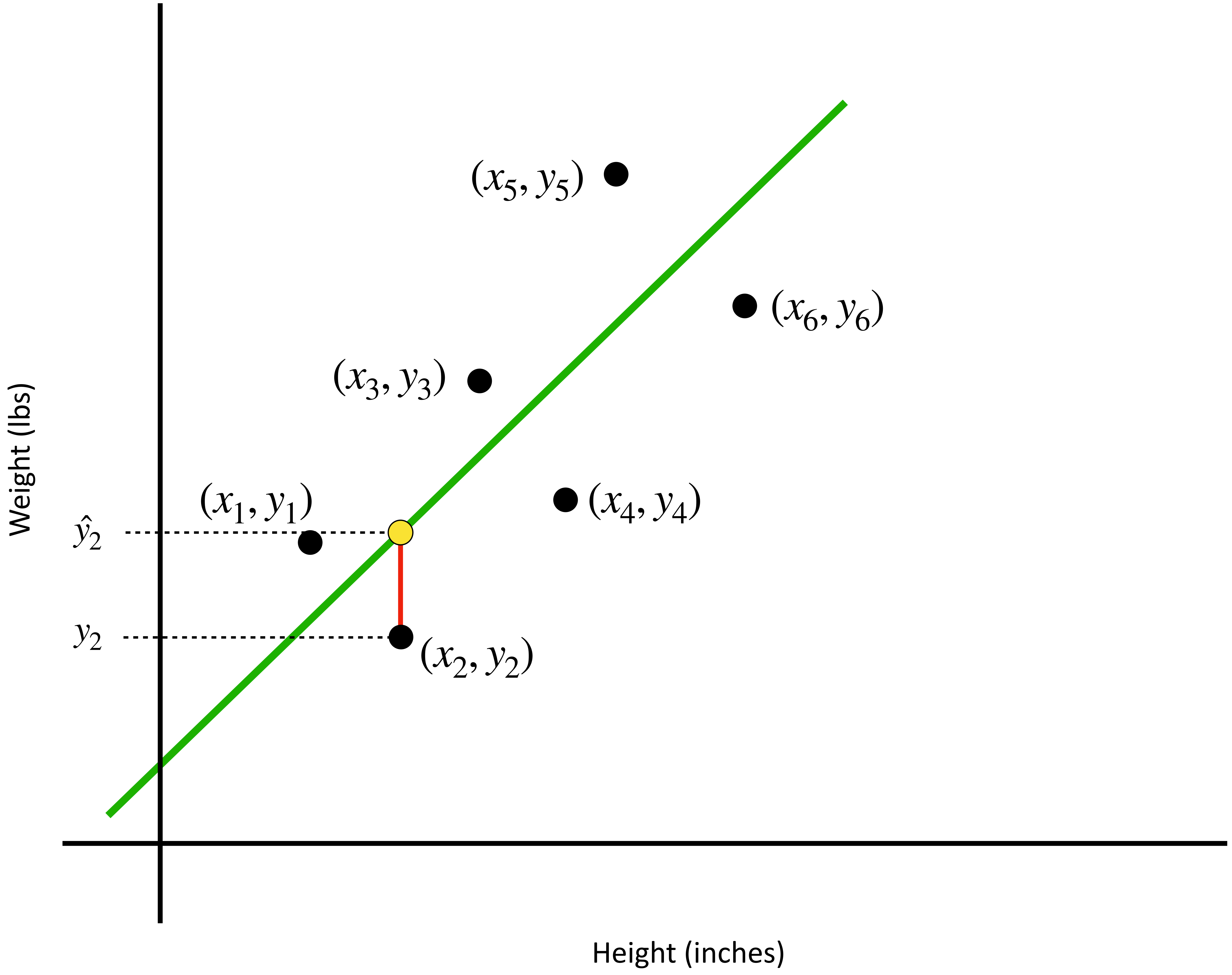
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)$$



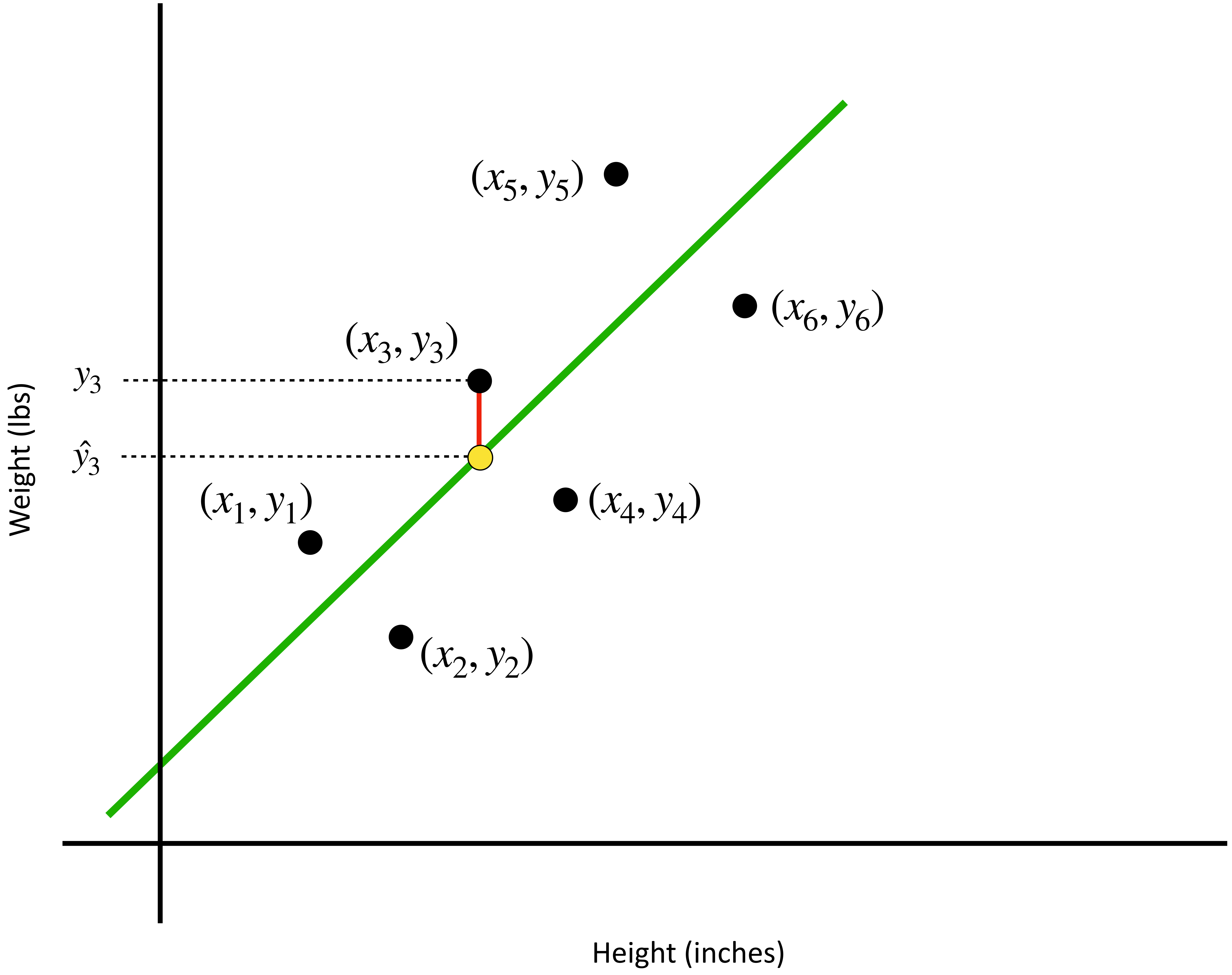
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$



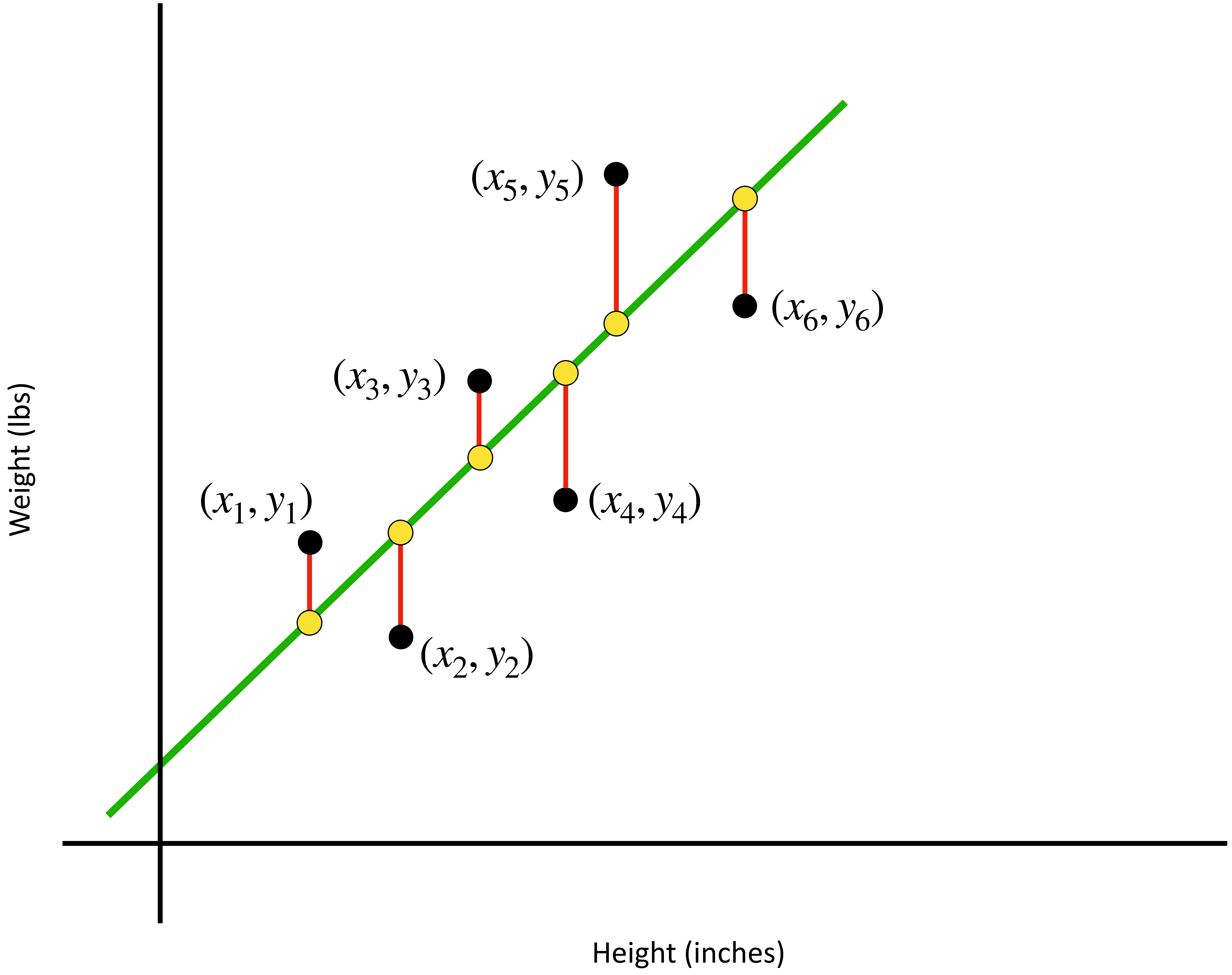
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$



Data is synthetic and not plotted to scale

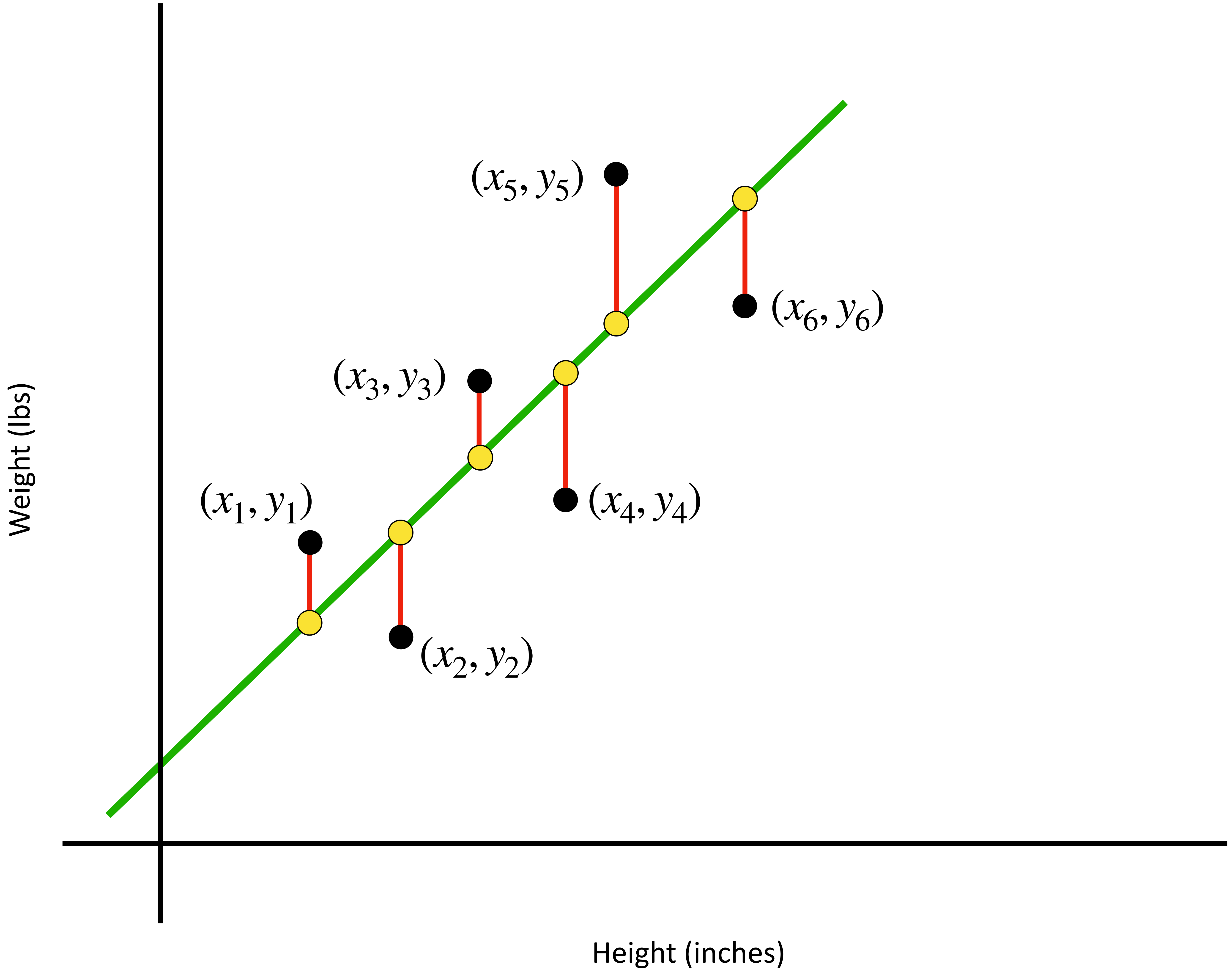
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$

But there's a small problem…



Data is synthetic and not plotted to scale

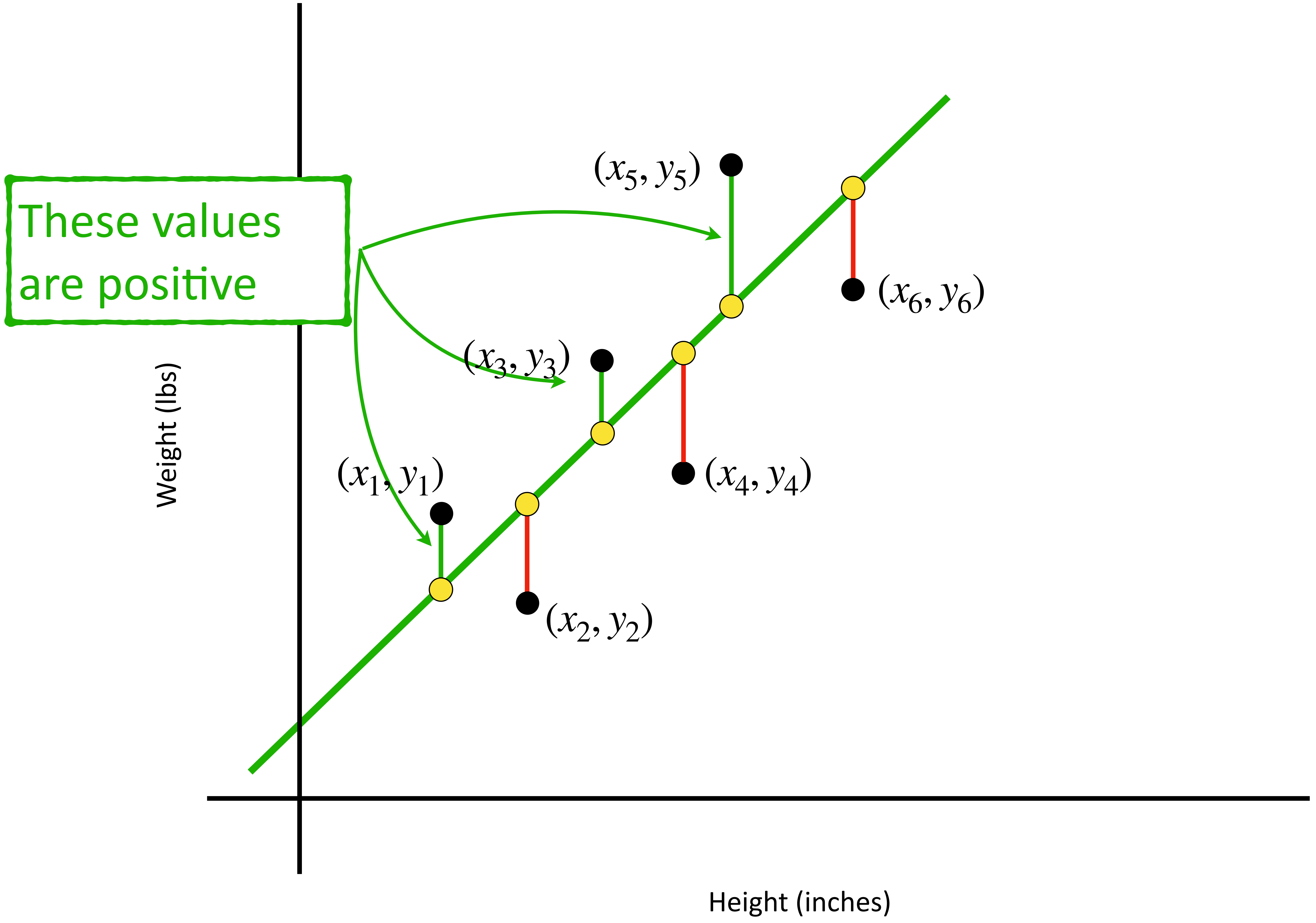
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$

But there's a small problem...



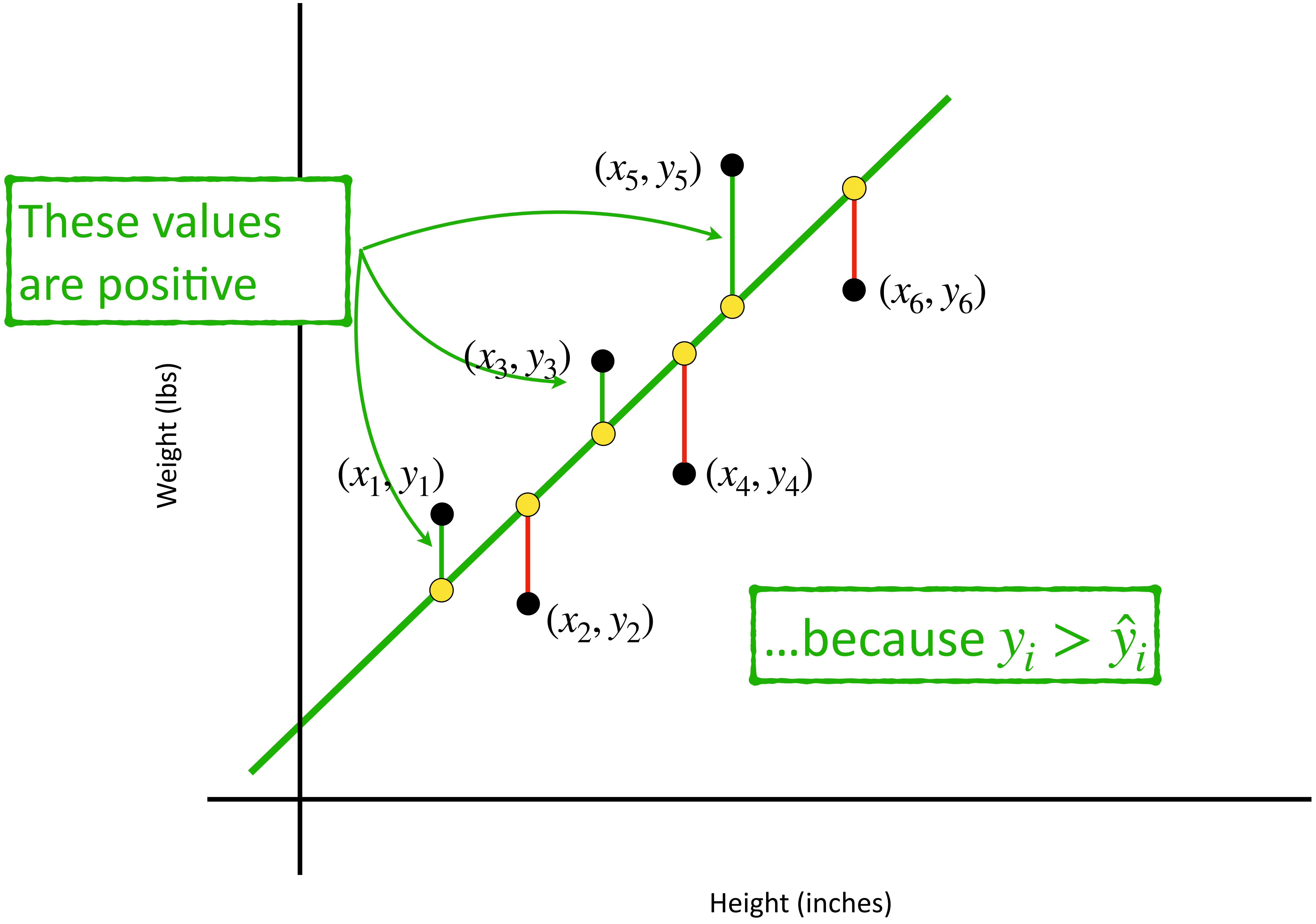
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$



But there's a small problem...

Data is synthetic and not plotted to scale

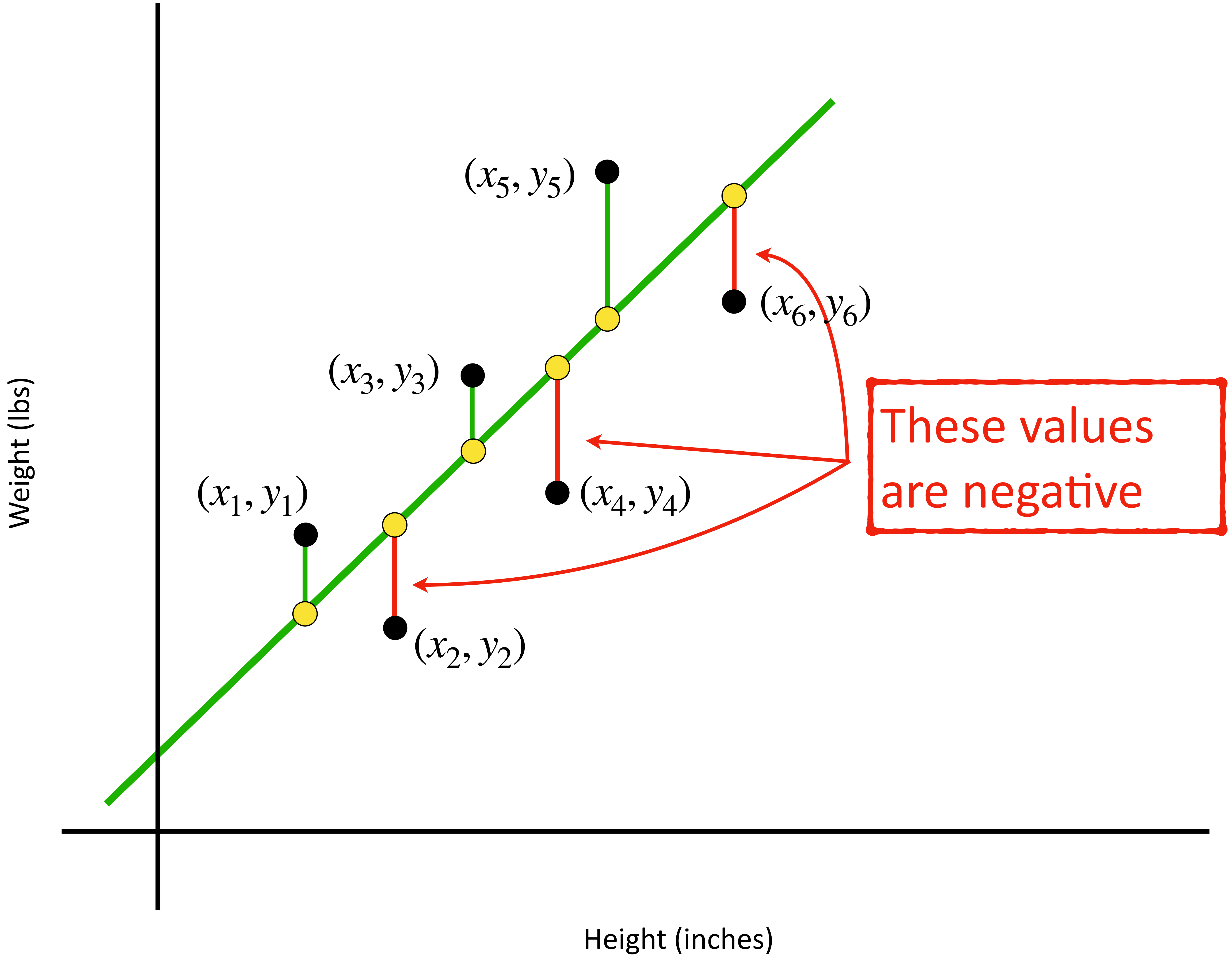
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+ (y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$

But there's a small problem…



Data is synthetic and not plotted to scale

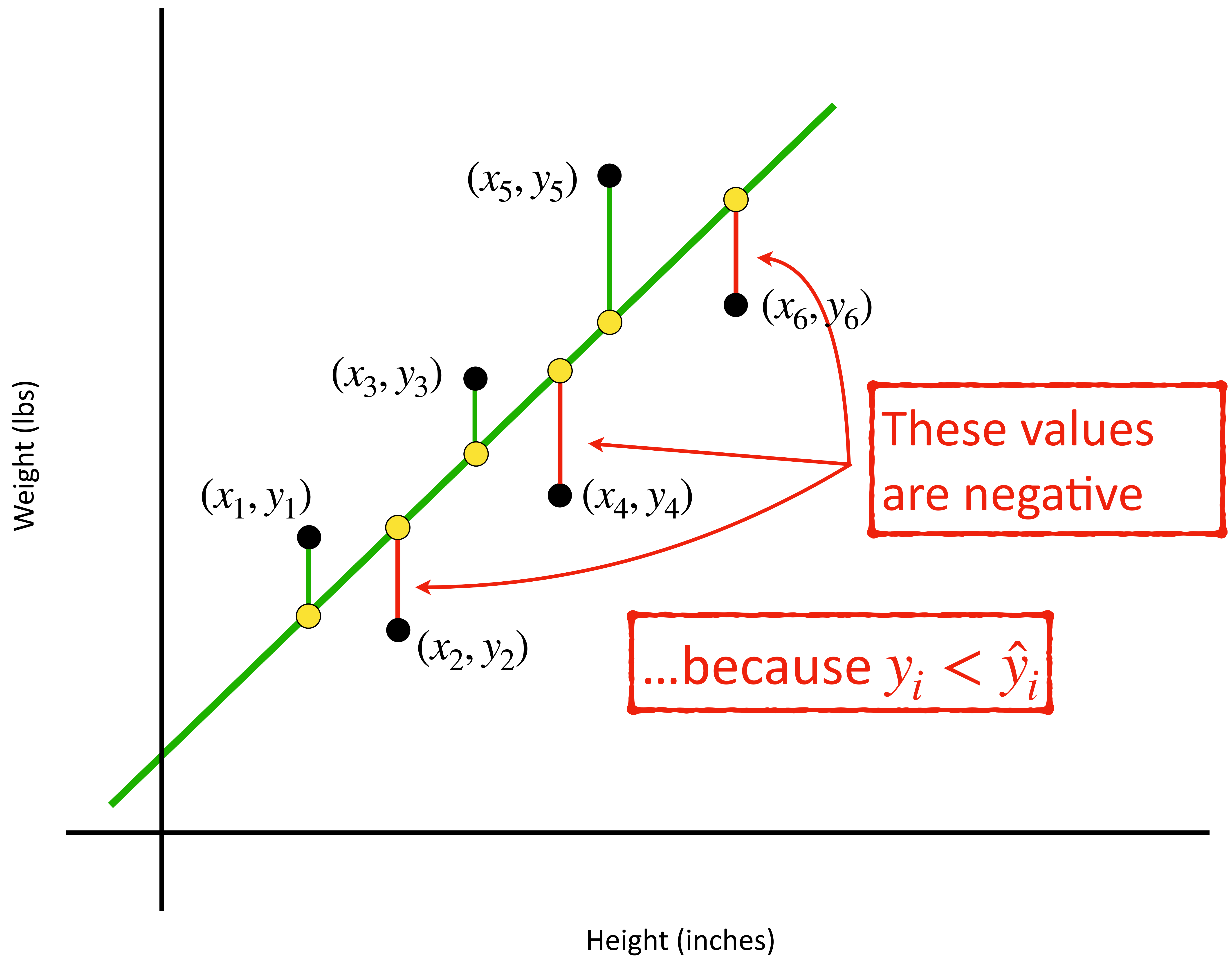
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$

But there's a small problem…



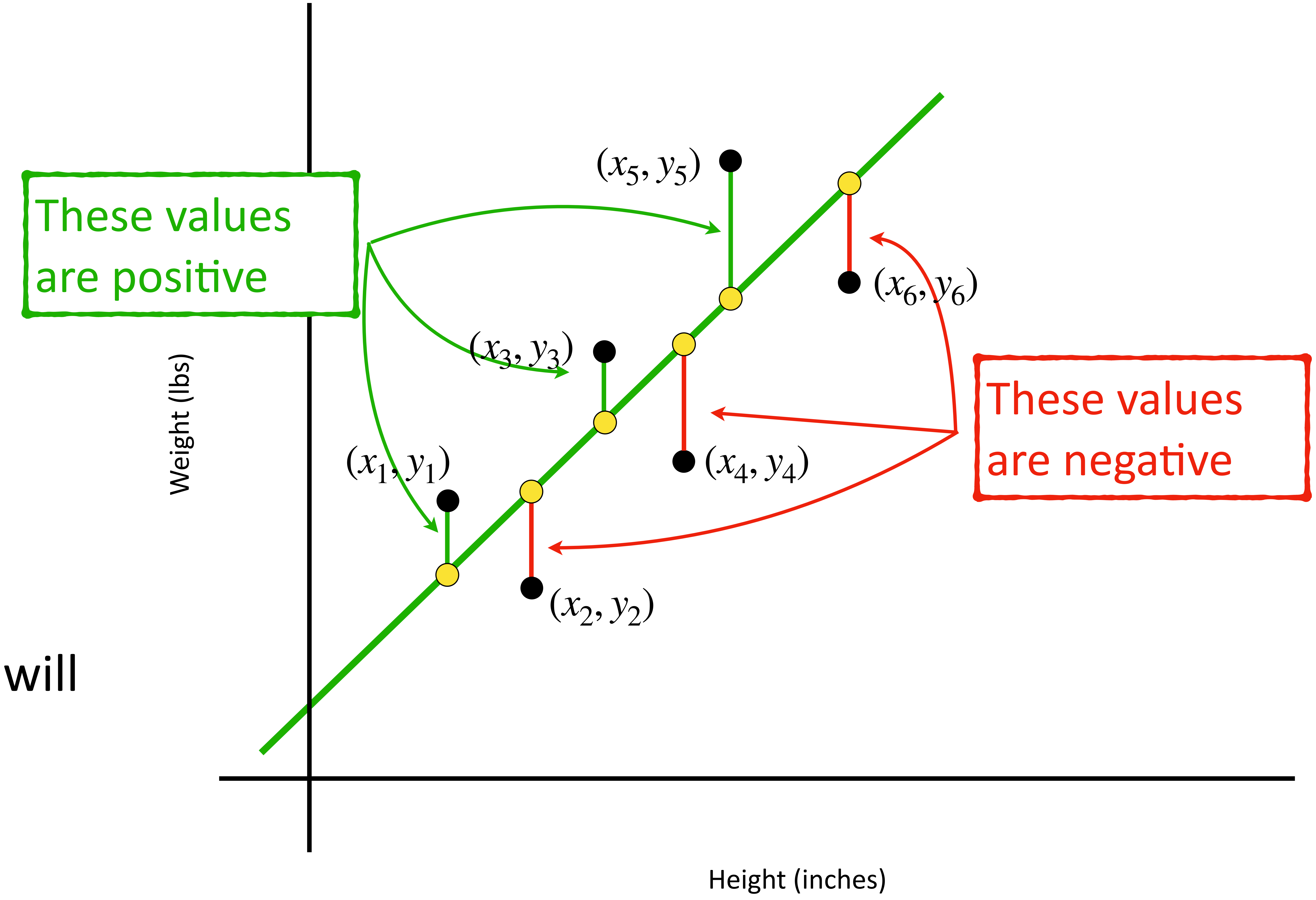
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3)$$
$$+(y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$



… the positive and negative values will cancel each other out

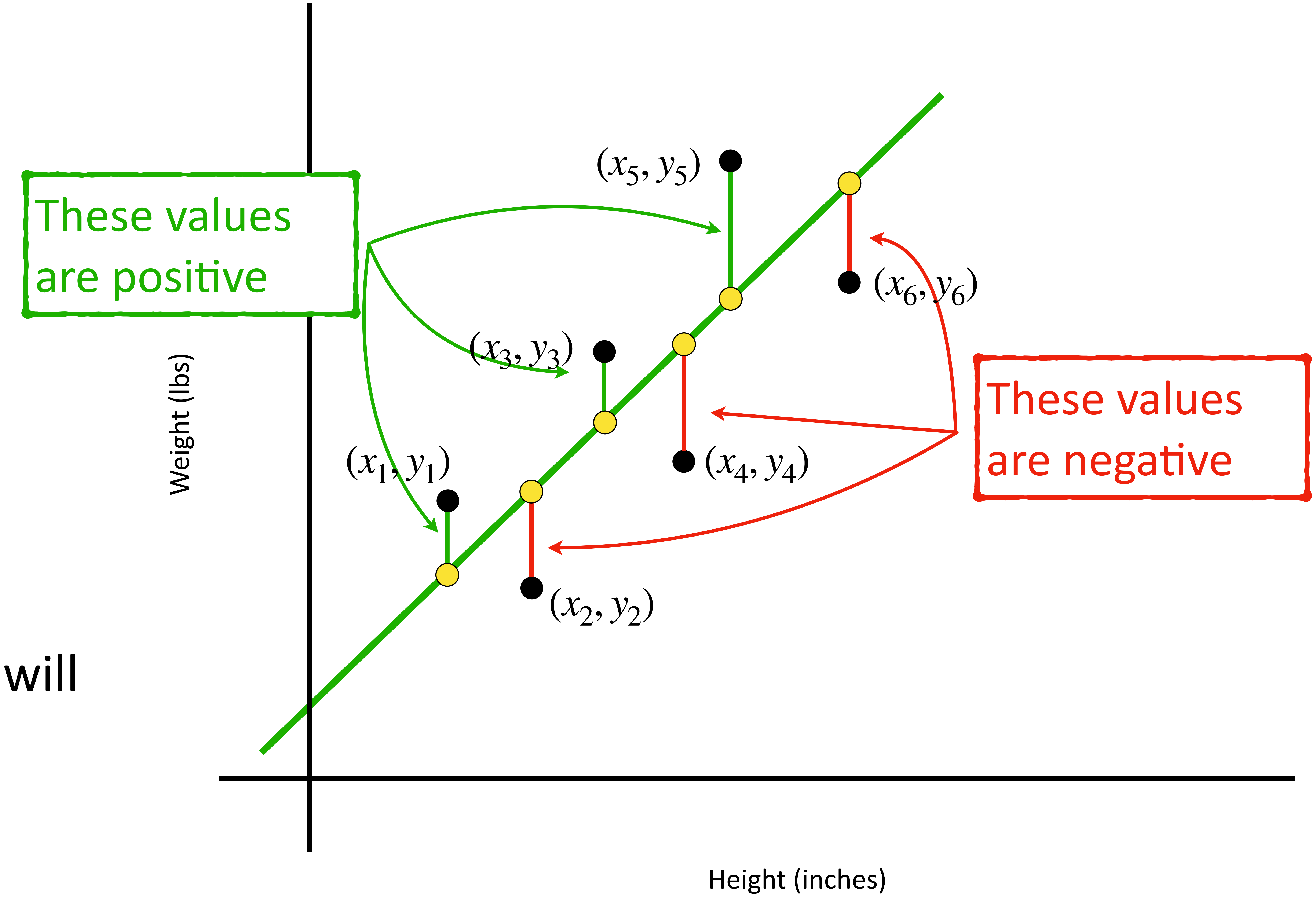
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

For each point calculate the distance to the line (the error)

$$(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2) + (y_3 - \hat{y}_3) + (y_4 - \hat{y}_4) + (y_5 - \hat{y}_5) + (y_6 - \hat{y}_6)$$



To fix that, we **square** each term

... the positive and negative values will cancel each other out

Data is synthetic and not plotted to scale

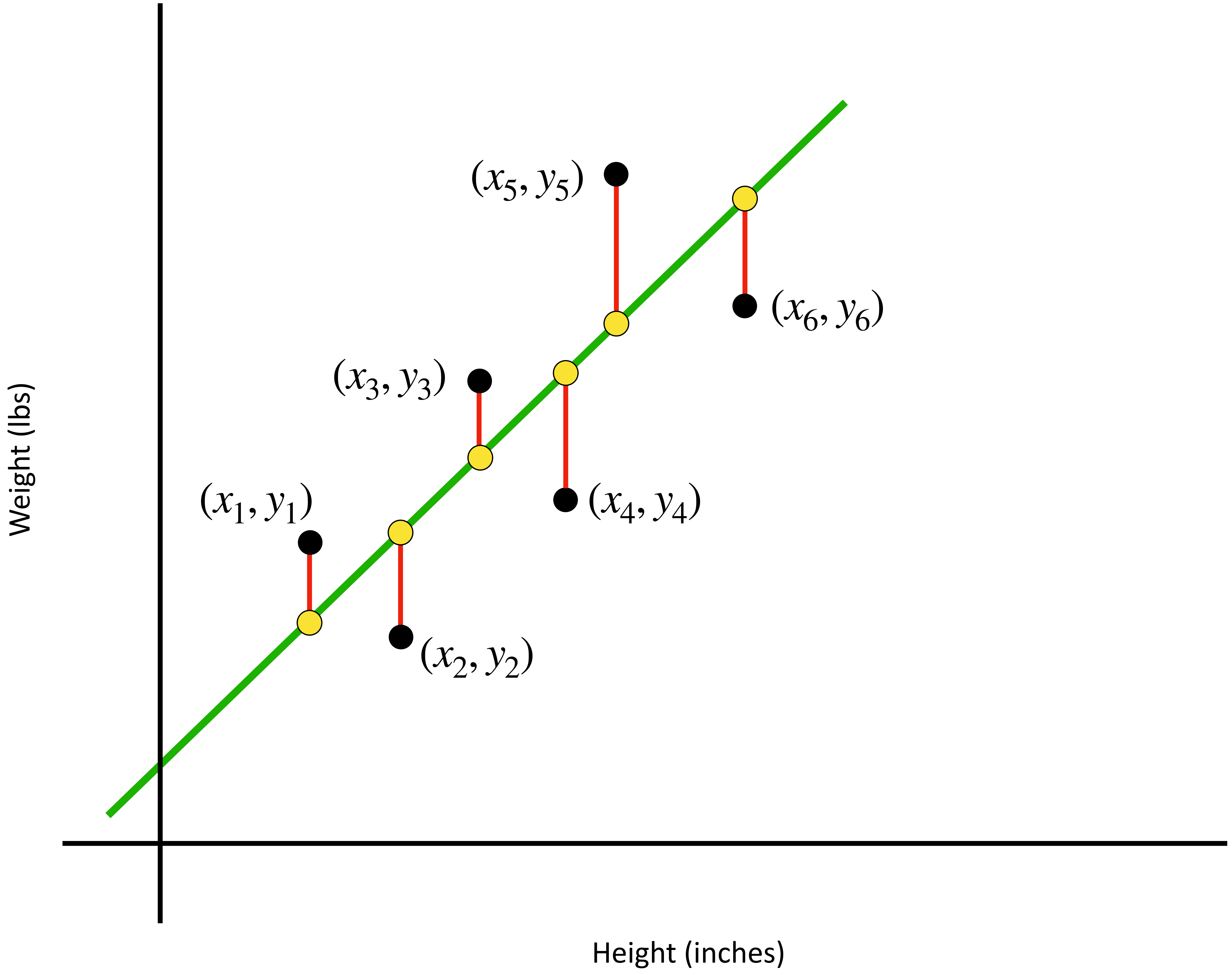
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2$$
$$+(y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2$$

To fix that, we **square** each term

This is the sum of squared errors.

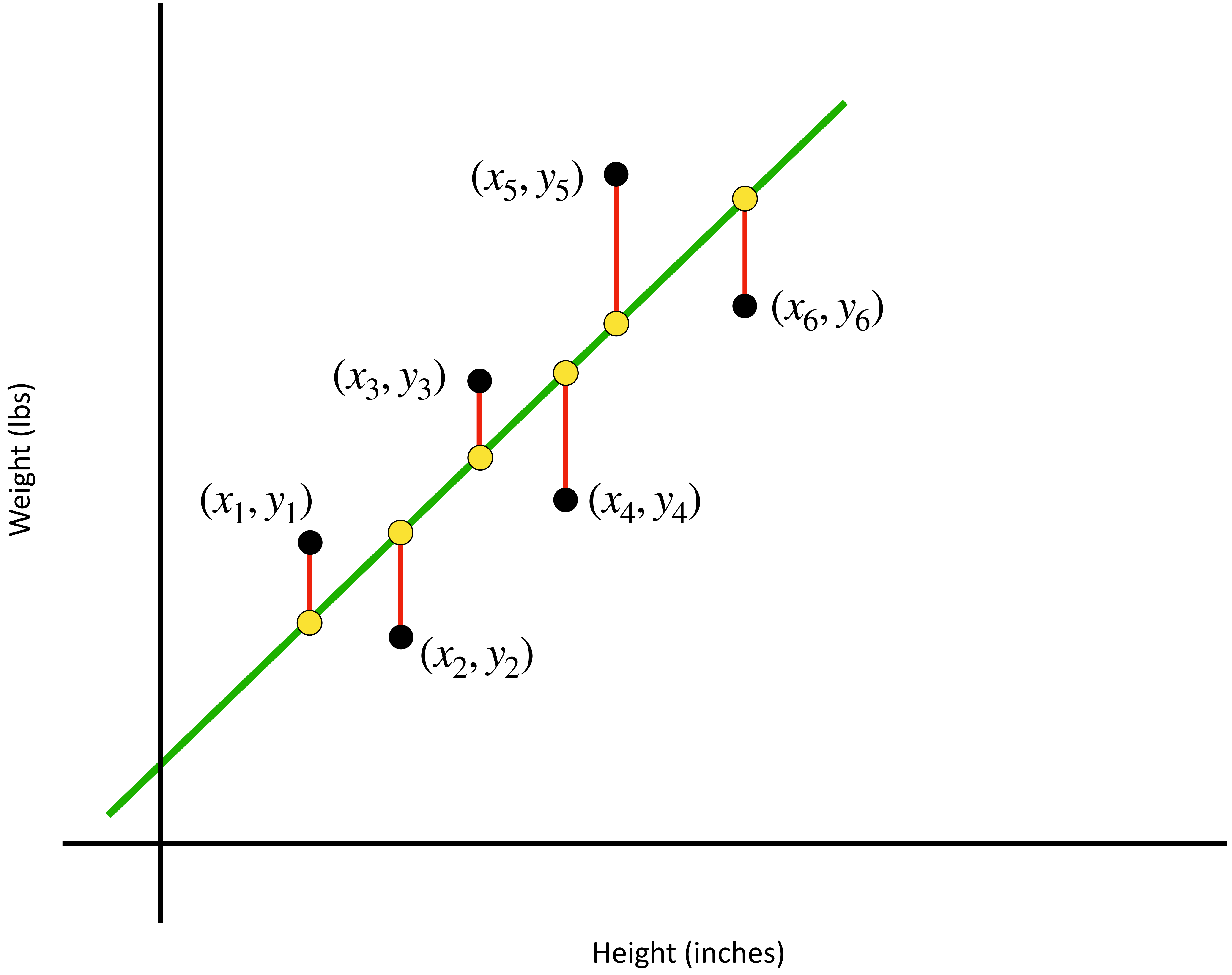


Data is synthetic and not plotted to scale

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

Sum of squared errors:

$$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2$$
$$+(y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2$$

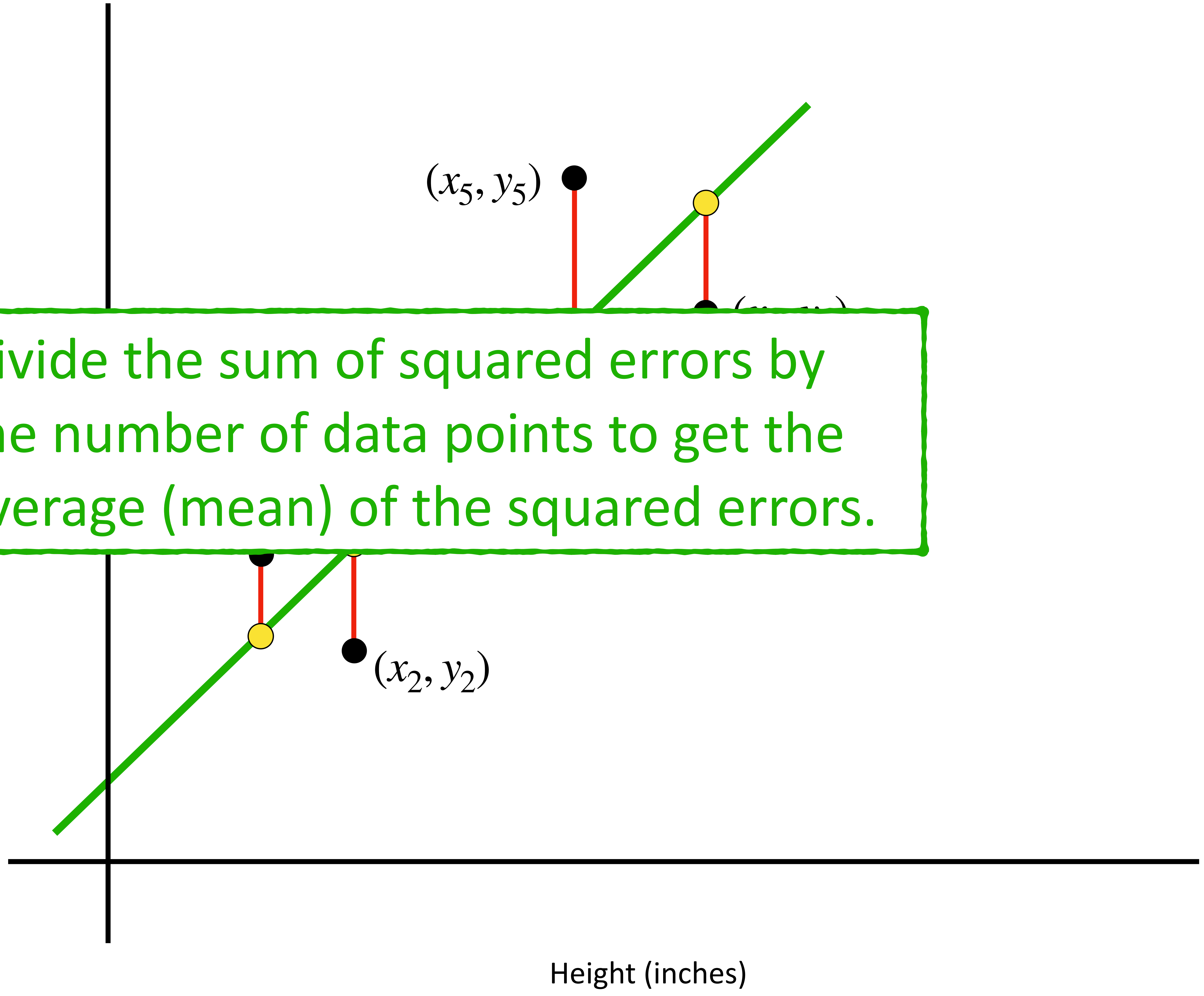


Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$



Divide the sum of squared errors by the number of data points to get the average (mean) of the squared errors.

Data is synthetic and not plotted to scale

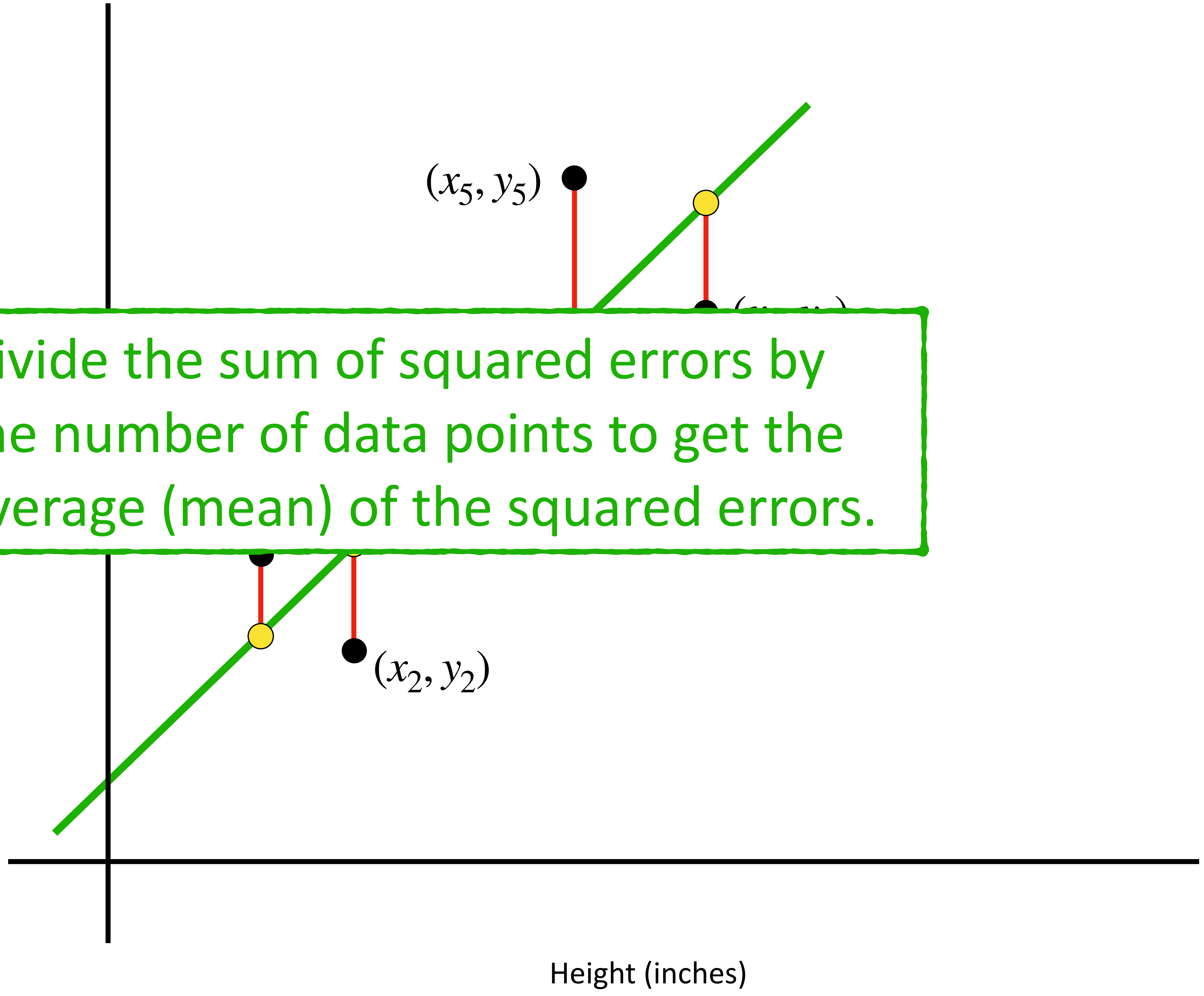
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



Divide the sum of squared errors by the number of data points to get the average (mean) of the squared errors.

Data is synthetic and not plotted to scale

Simple Linear Regression

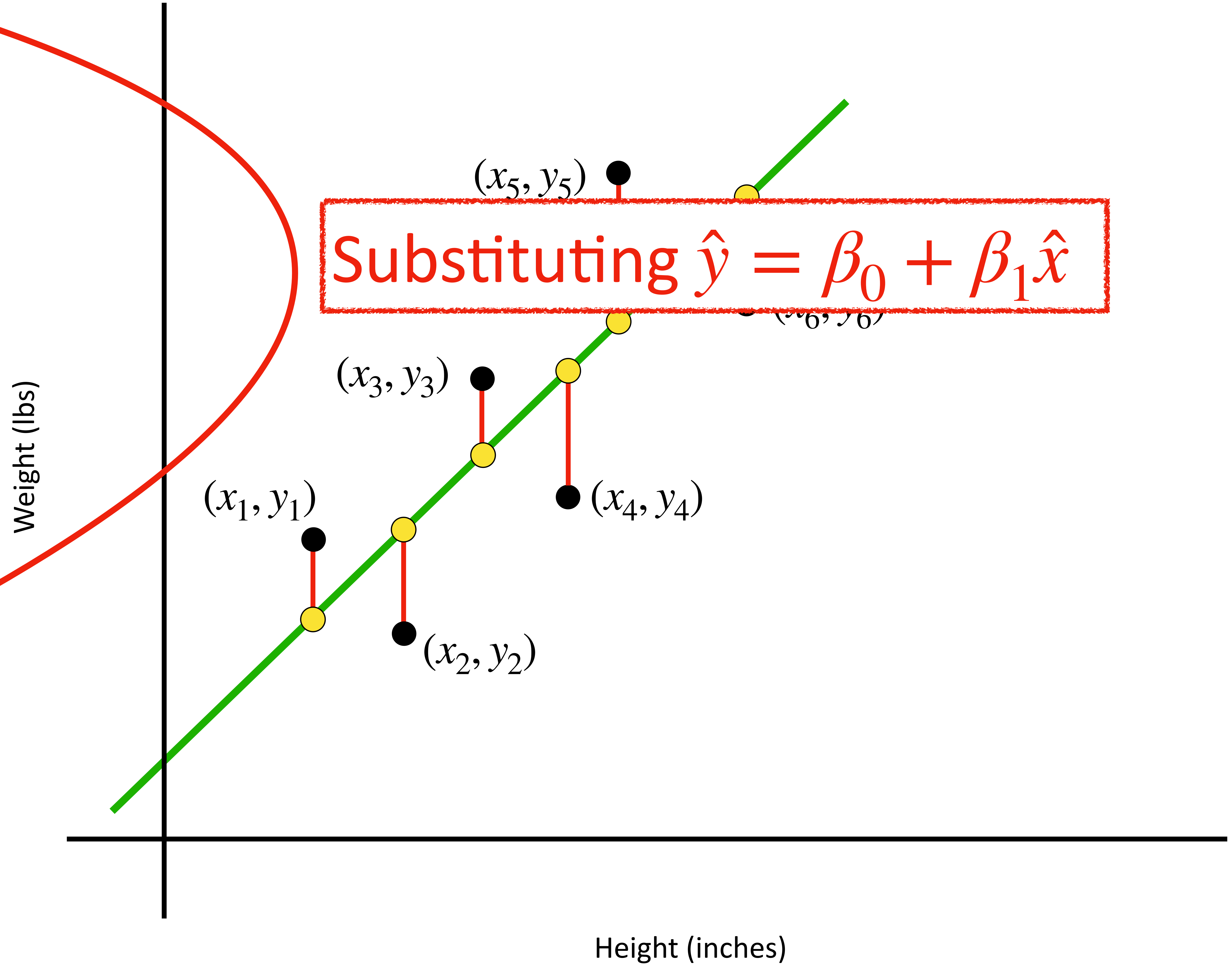
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$

Substituting $\hat{y} = \beta_0 + \beta_1 \hat{x}$



Data is synthetic and not plotted to scale

The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

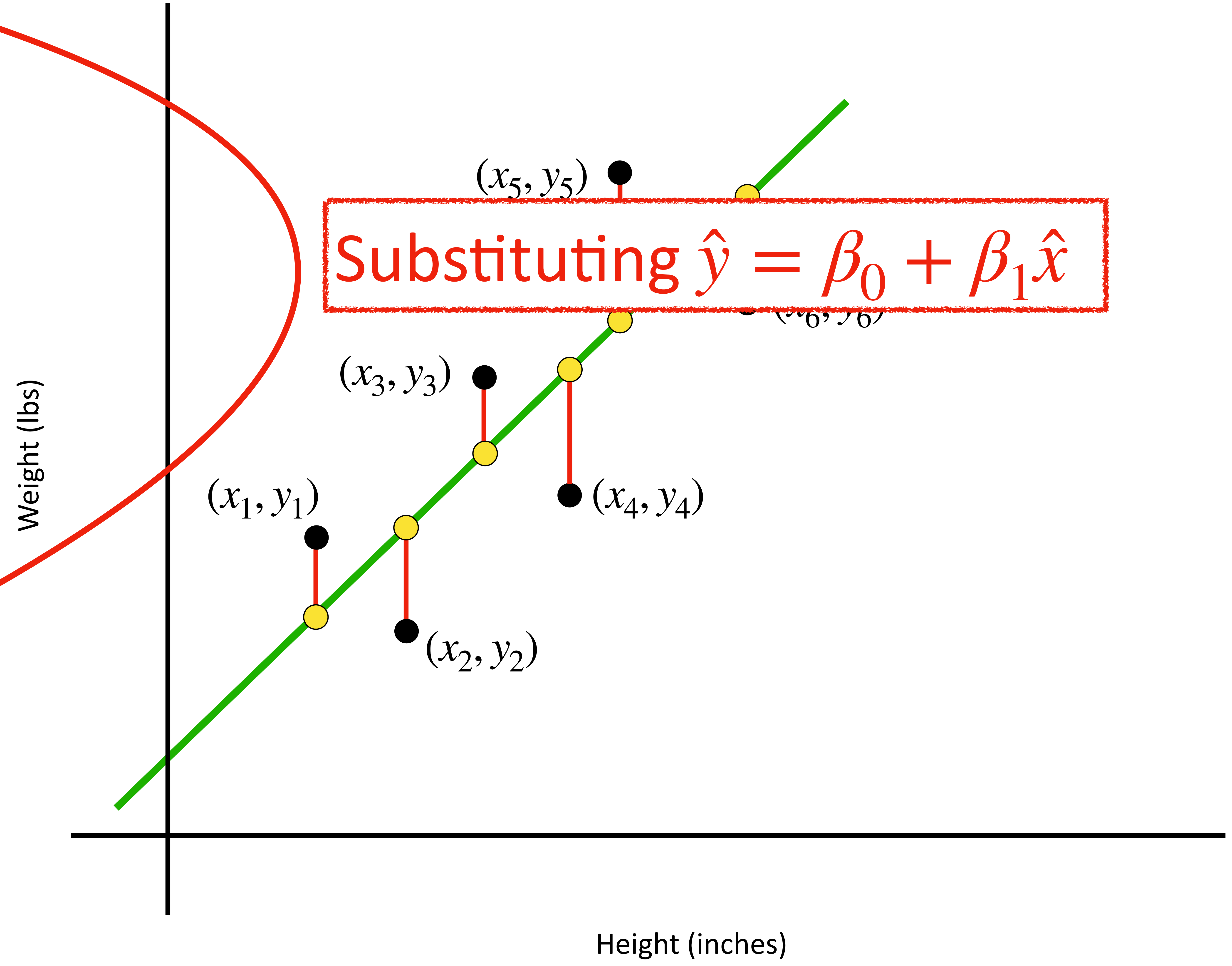
$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$

For every value of i, $\hat{x}_i$ equals x_i

Substituting $\hat{y} = \beta_0 + \beta_1 \hat{x}$



Data is synthetic and not plotted to scale

Simple Linear Regression

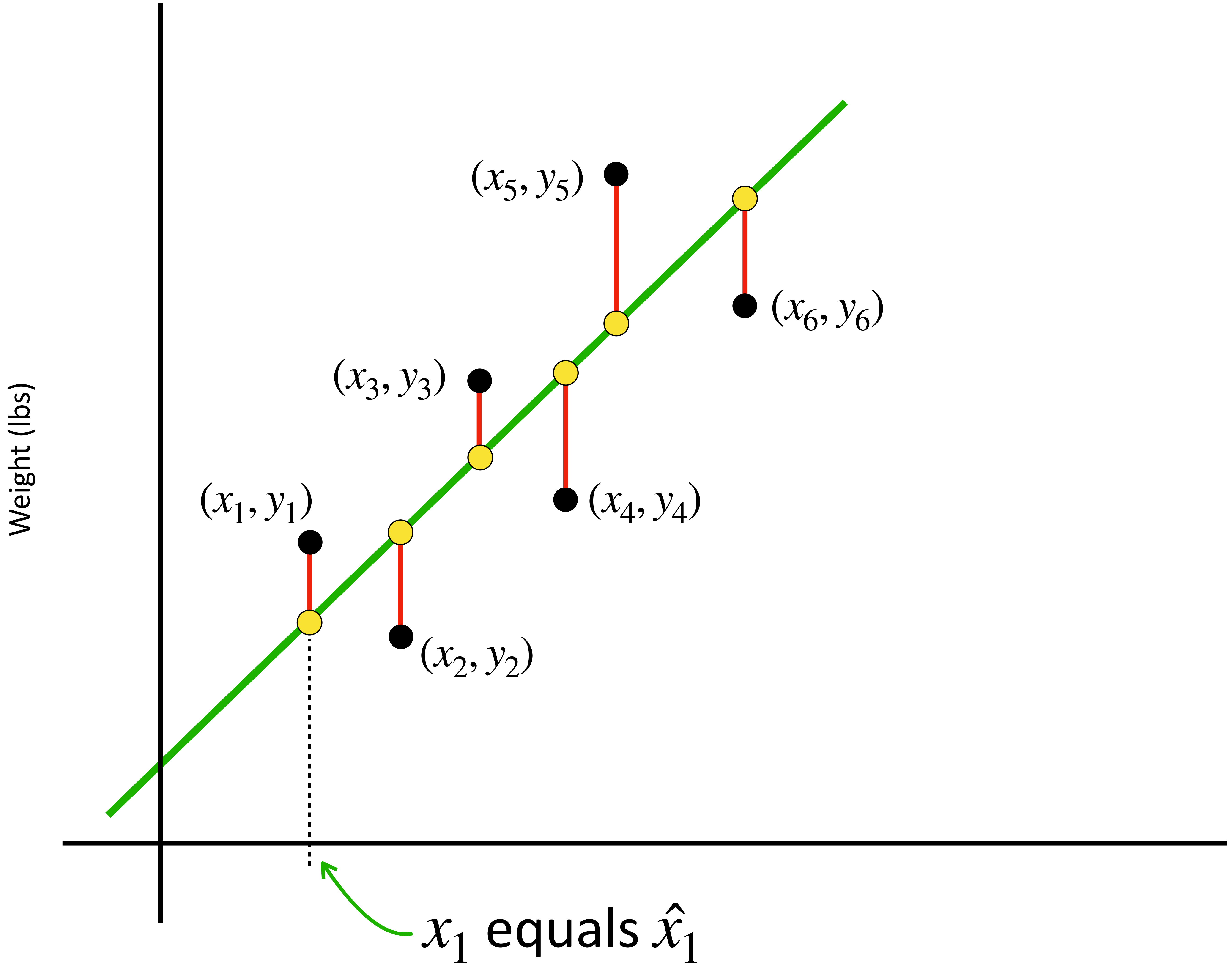
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{\begin{array}{c}(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 \\ +(y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2\end{array}}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$

$\boxed{\text{For every value of } i, \hat{x}_i \text{ equals } x_i}$



Data is synthetic and not plotted to scale

Simple Linear Regression

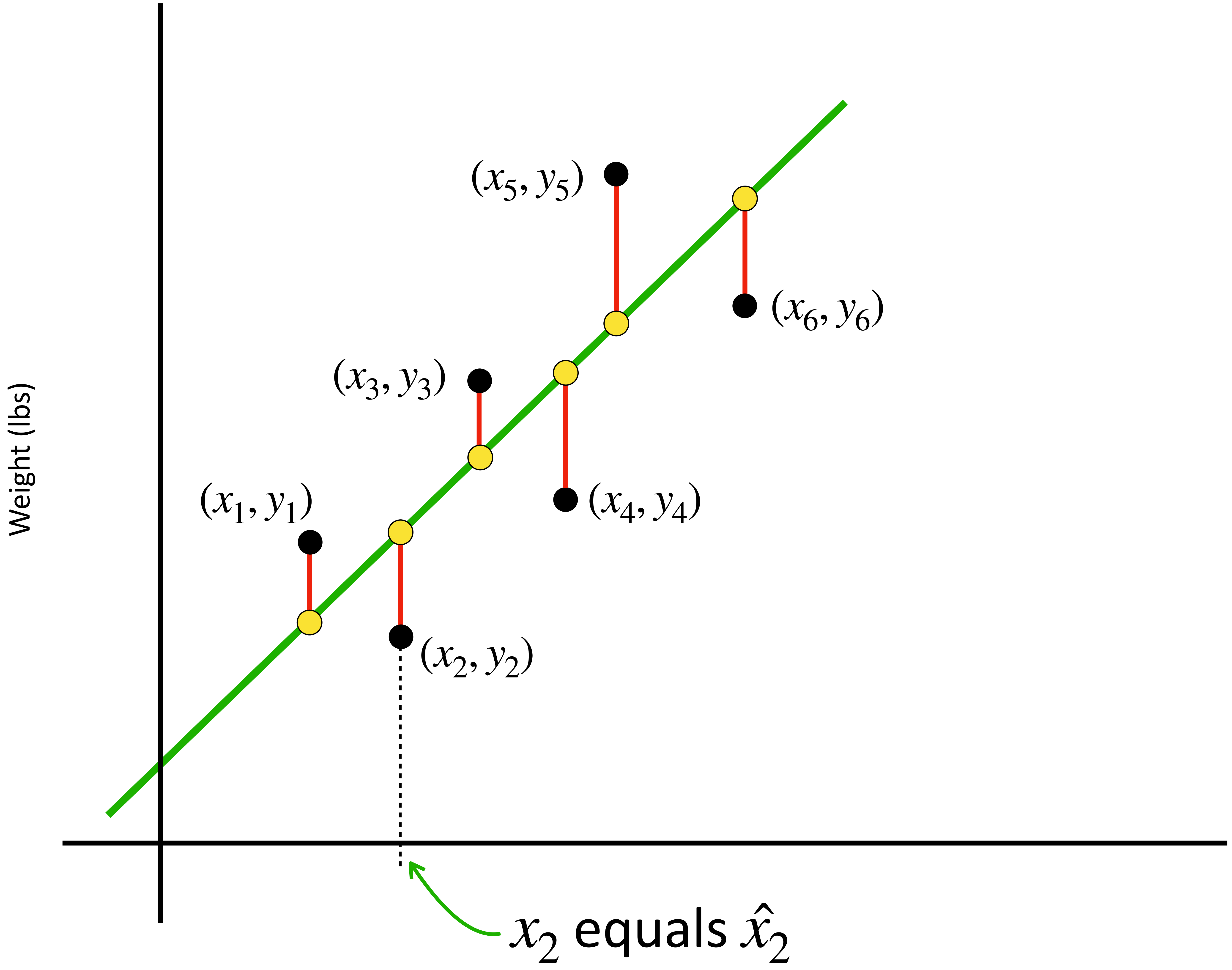
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1\hat{x}_i)^2$$

For every value of i, $\hat{x}_i$ equals x_i



Data is synthetic and not plotted to scale

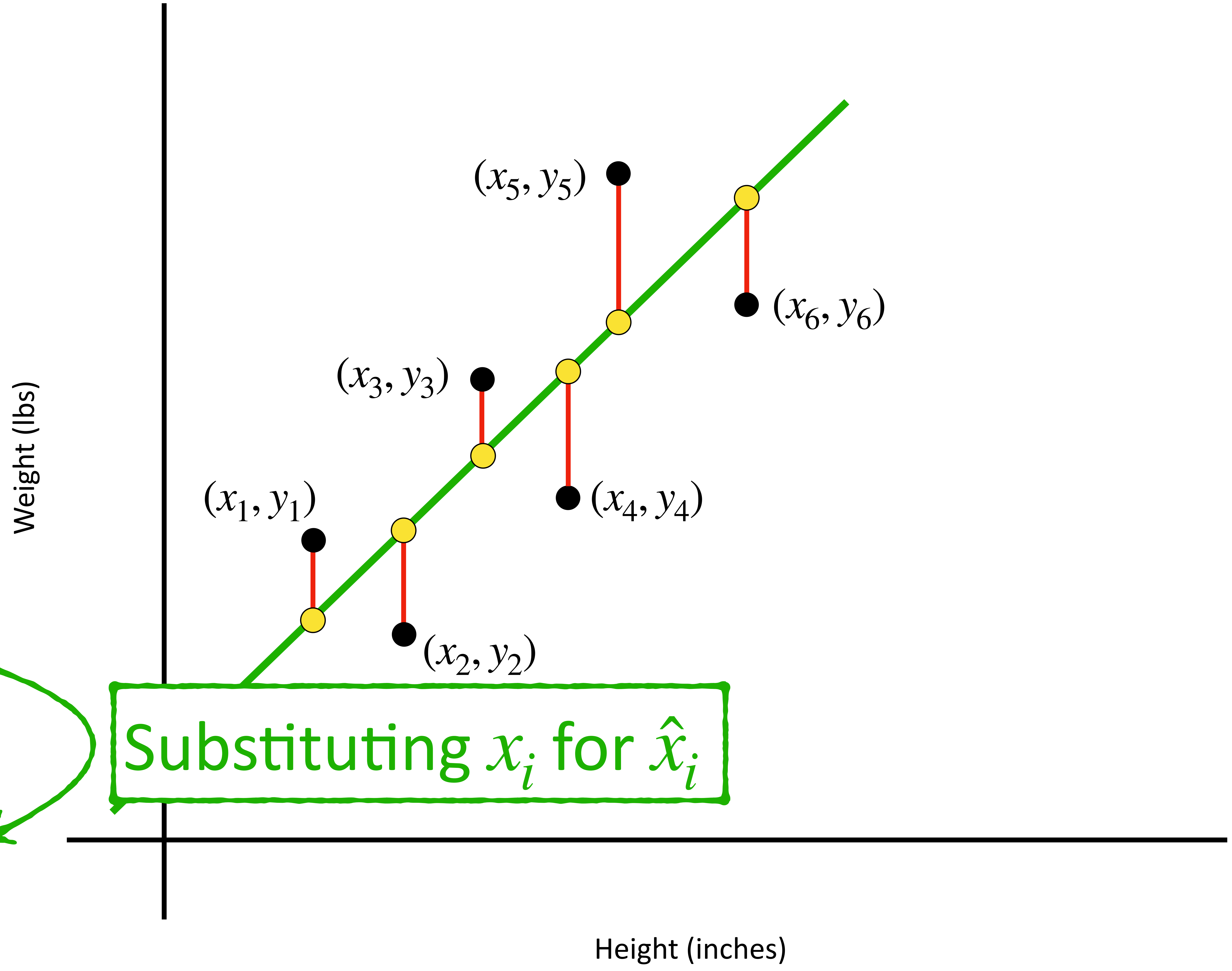
The line of best fit is $\hat{y} = \beta_0 + \beta_1 \hat{x}$

$$\frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2 + (y_5 - \hat{y}_5)^2 + (y_6 - \hat{y}_6)^2}{n}$$

This is the **Mean Square Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 \hat{x}_i)^2$$

$$= \frac{1}{n}\sum_{i=0}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$



Data is synthetic and not plotted to scale

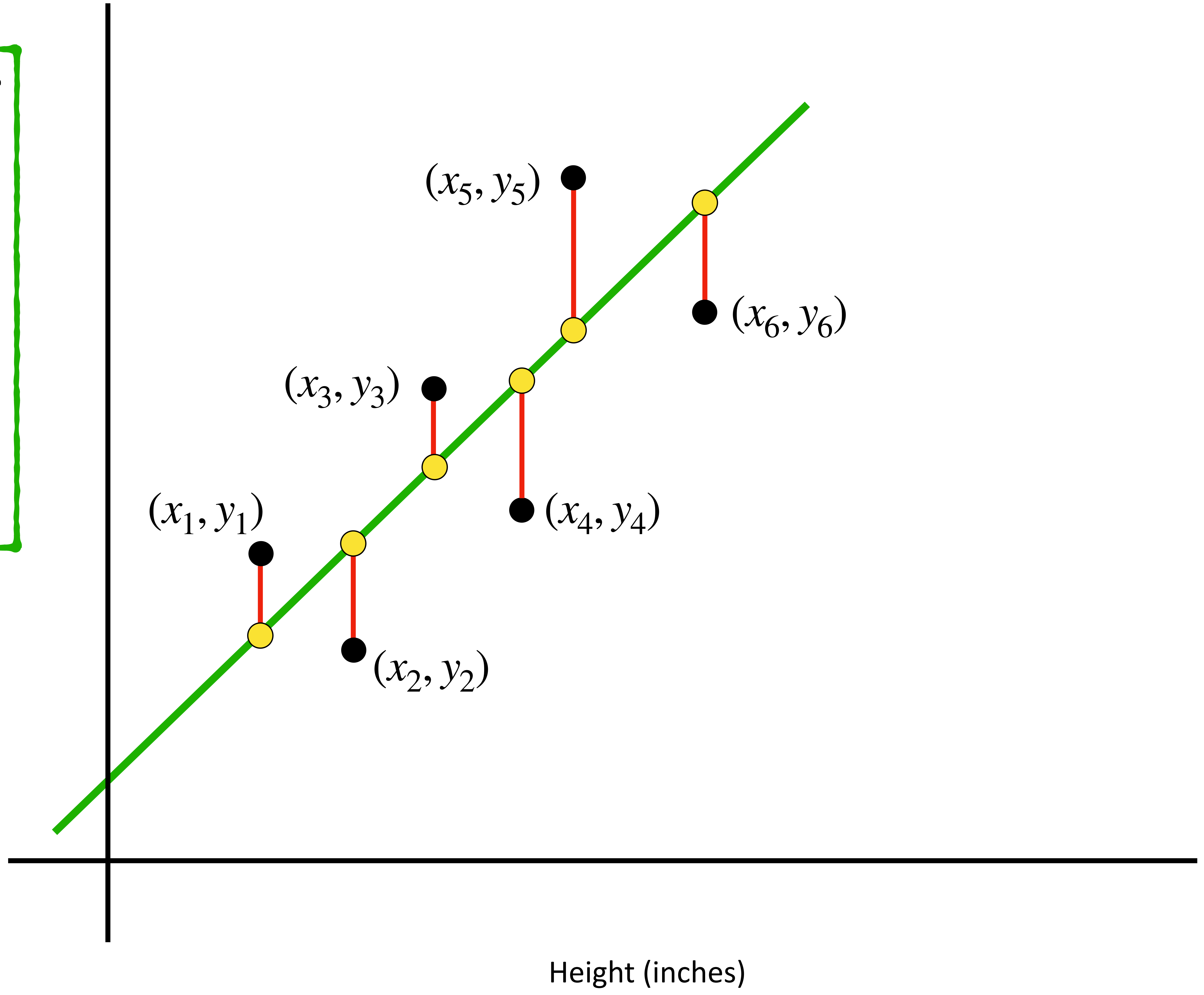
Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 x$

Problem Statement: Given a set of data points in $\mathbf{R}^2$,

$(x_1, y_1), (x_2, y_2), (x_3, y_3) \ldots (x_n, y_n)$,

find the line that minimizes the **Mean Squared Error (MSE)**



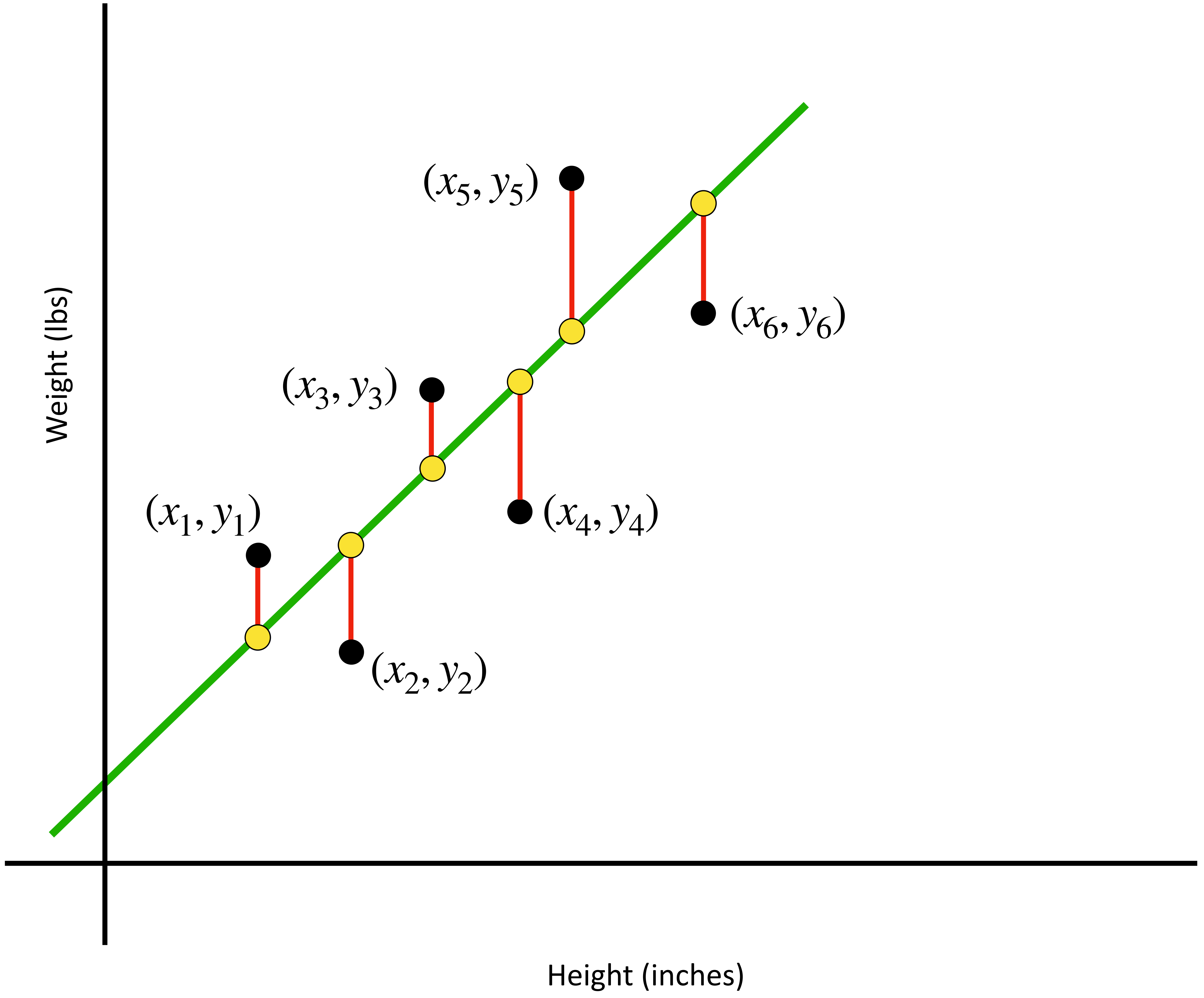
Data is synthetic and not plotted to scale

Simple Linear Regression

The line of best fit is $\hat{y} = \beta_0 + \beta_1 x$

This is the **Mean Squared Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$



Data is synthetic and not plotted to scale

Simple Linear Regression

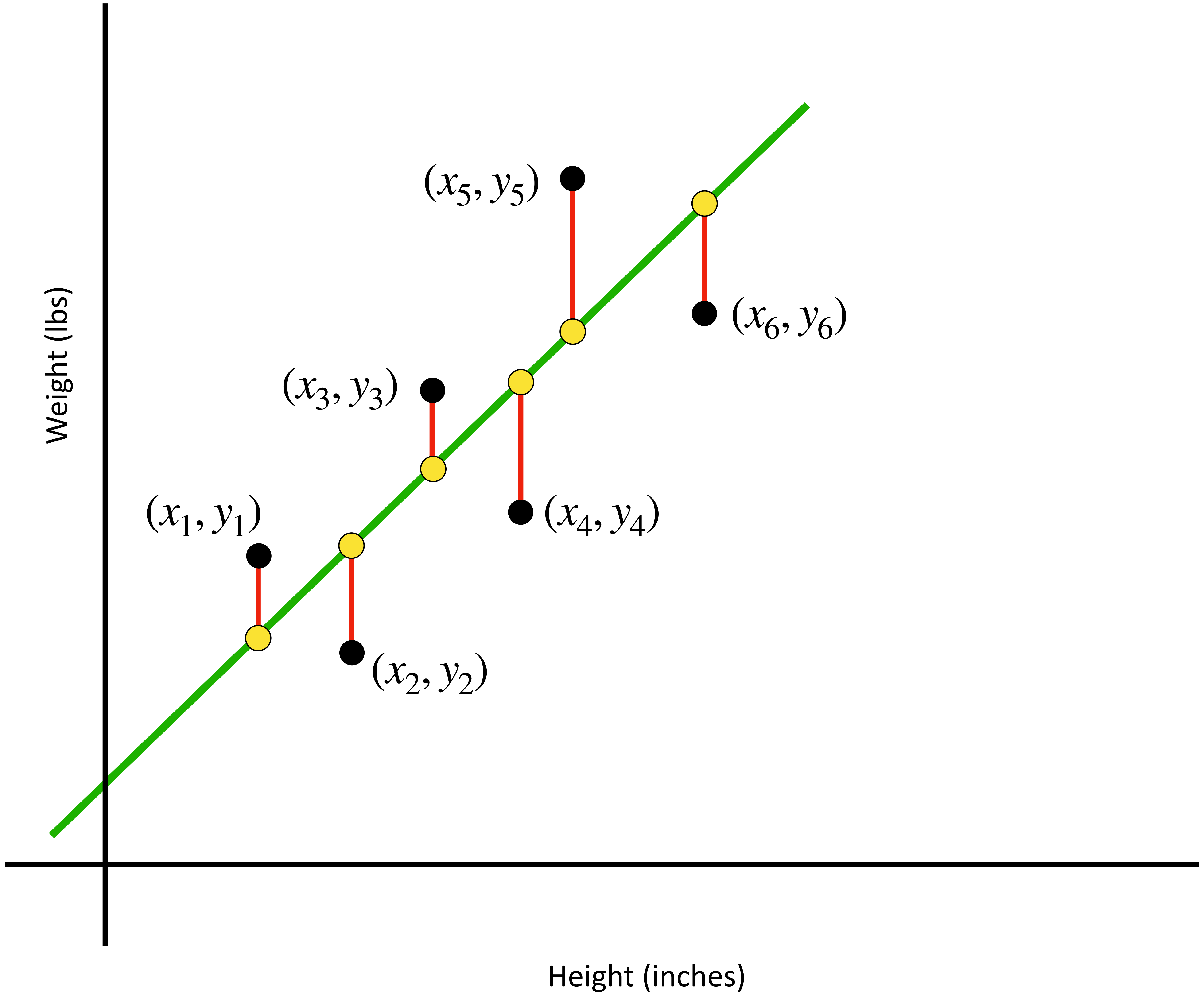
The line of best fit is $\hat{y} = \beta_0 + \beta_1 x$

This is the **Mean Squared Error (MSE)**

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

The Problem Statement:

Simple Linear Regression: Find the values of β_0 and β_1 such that the Mean Squared Error (MSE) is minimized.

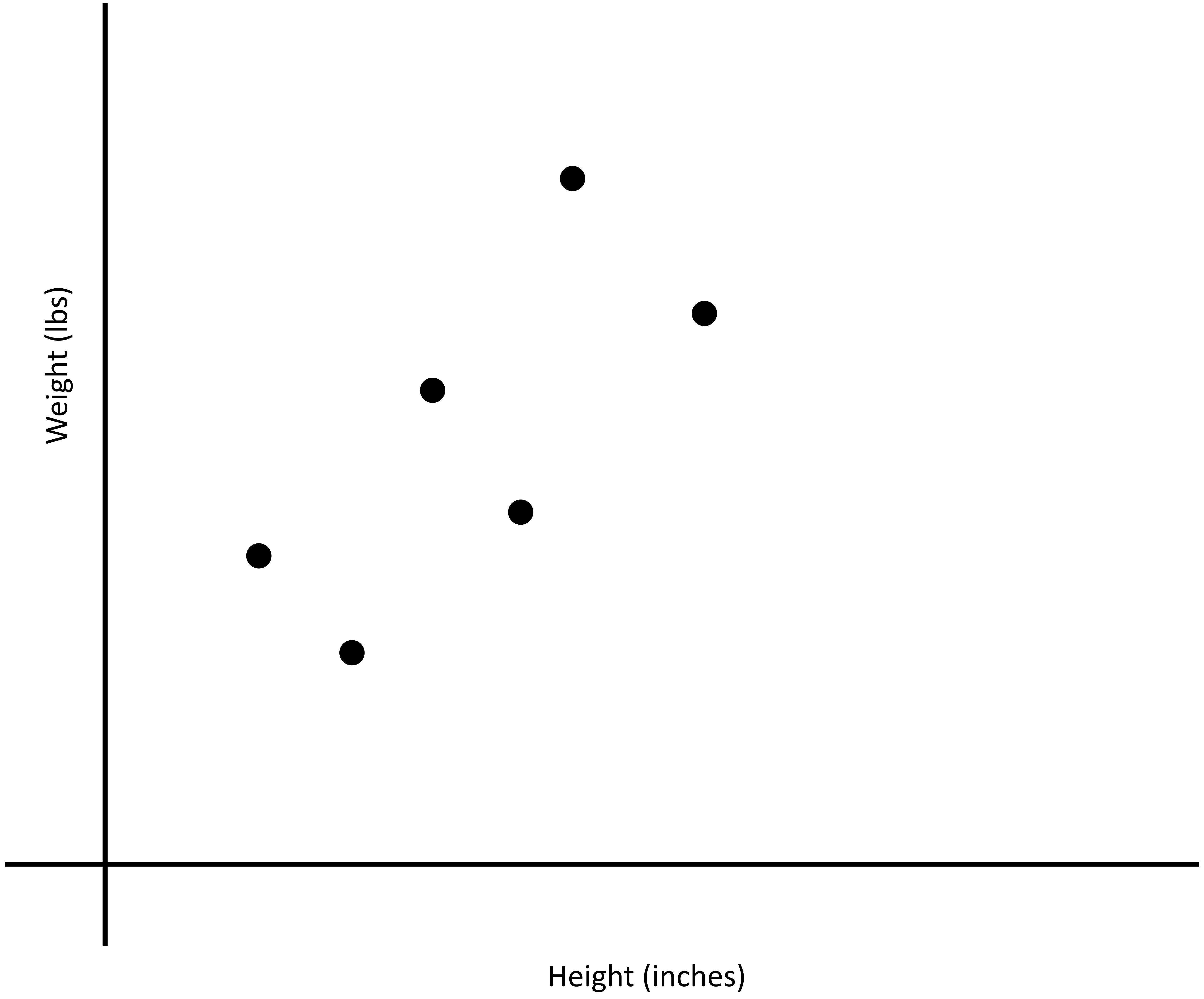


Data is synthetic and not plotted to scale

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$



The Problem Statement:

Simple Linear Regression: Find the values of β_0 and β_1 such that the Mean Squared Error (MSE) is minimized.

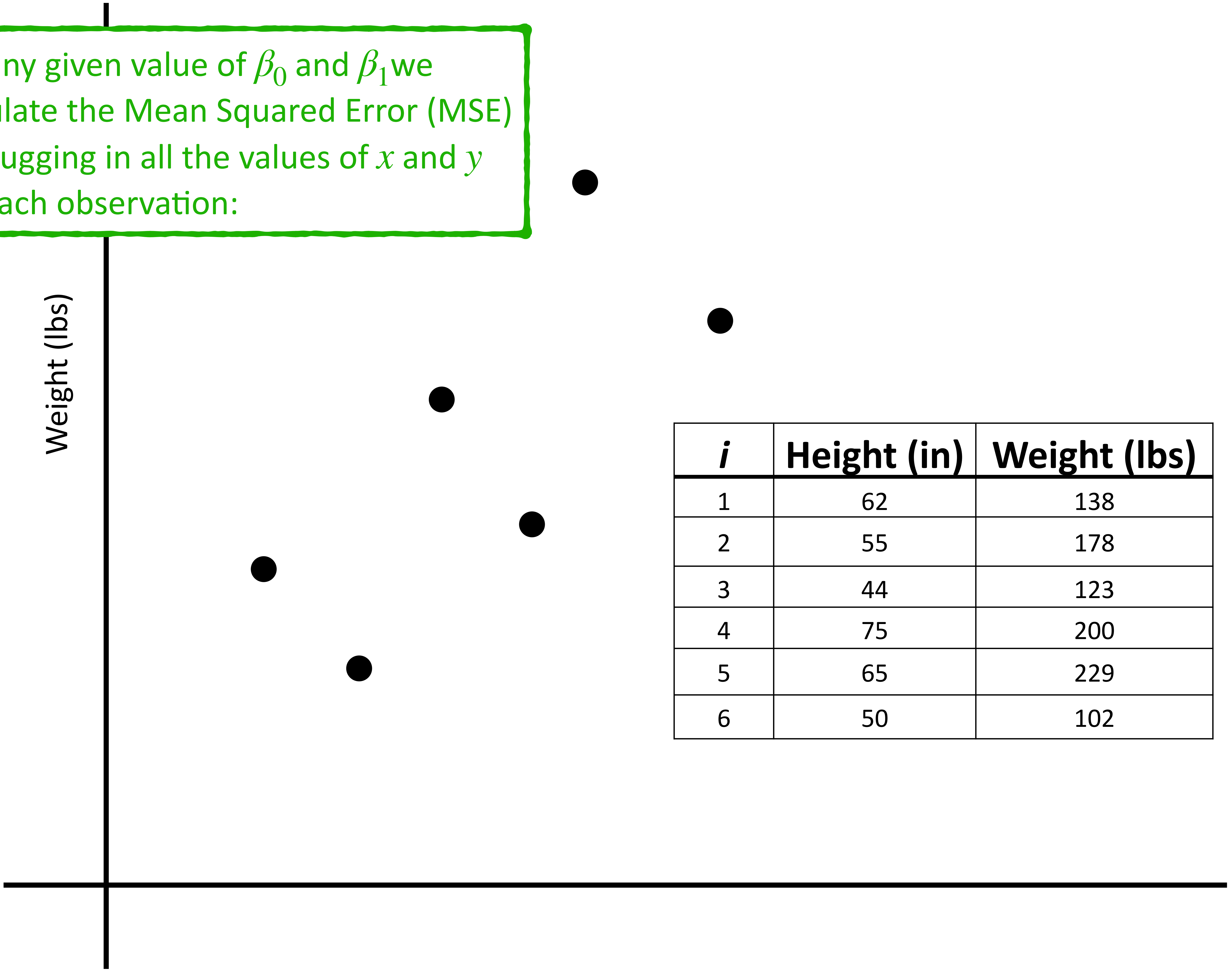
Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{(62-\beta_0-\beta_1 138)^2+(55-\beta_0-\beta_1 178)^2+(44-\beta_0-\beta_1 123)^2+(75-\beta_0-\beta_1 200)^2+(65-\beta_0-\beta_1 229)^2+(50-\beta_0-\beta_1 102)^2}{6}$$

Simple Linear Regression

For any given value of β_0 and β_1 we calculate the Mean Squared Error (MSE) by plugging in all the values of x and y for each observation:



i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

The Problem Statement:

Simple Linear Regression: Find the values of β_0 and β_1 such that the Mean Squared Error (MSE) is minimized.

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

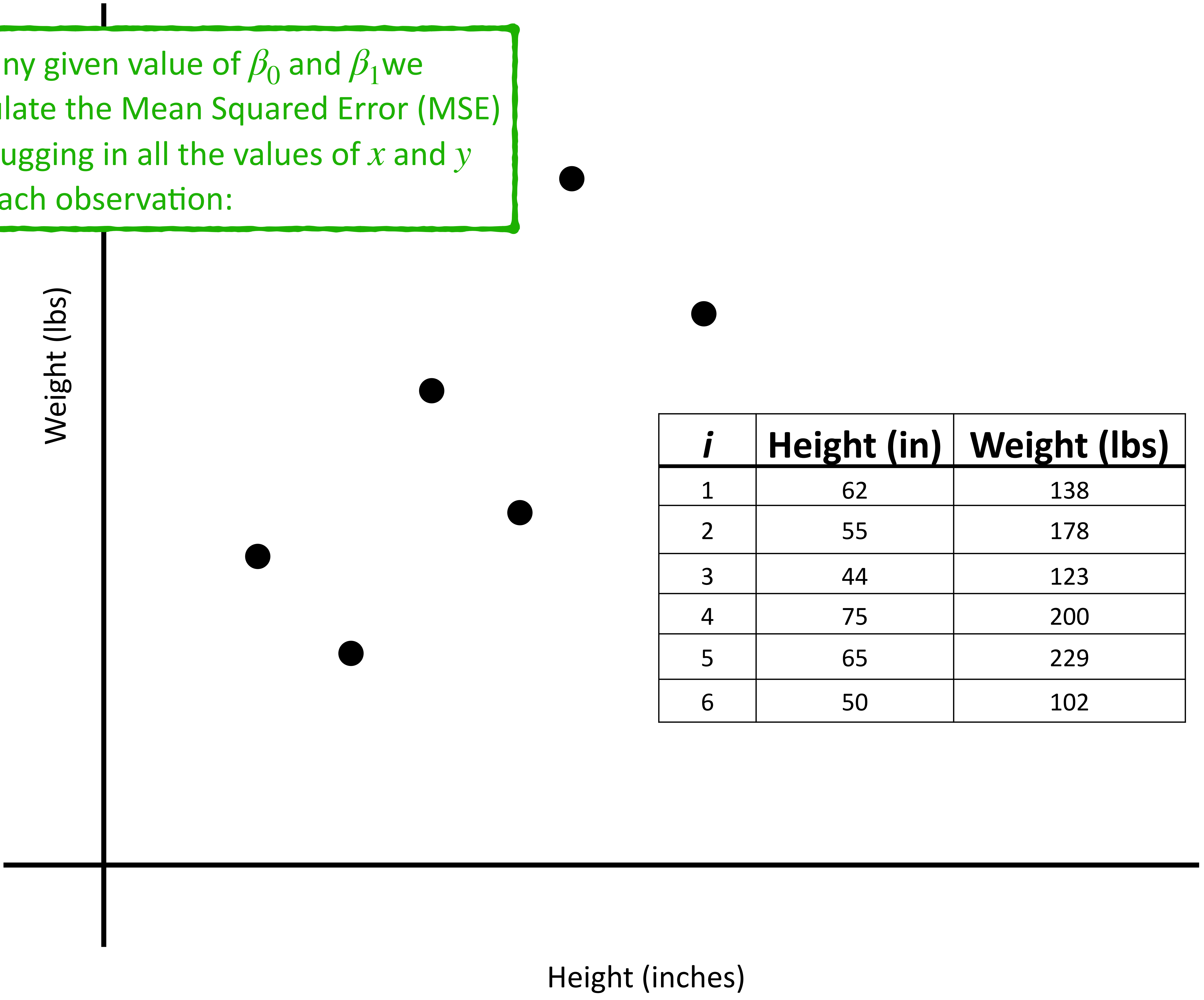
For any given value of β_0 and β_1 we calculate the Mean Squared Error (MSE) by plugging in all the values of x and y for each observation:

For example: If $\beta_0 = 3$ and $\beta_1 = -0.1$
then MSE = 430

$$\frac{\begin{array}{c}(62-\beta_0-\beta_1 138)^2 + (55-\beta_0-\beta_1 178)^2 + (44-\beta_0-\beta_1 123)^2 \\ +(75-\beta_0-\beta_1 200)^2 + (65-\beta_0-\beta_1 229)^2 + (50-\beta_0-\beta_1 102)^2\end{array}}{6}$$

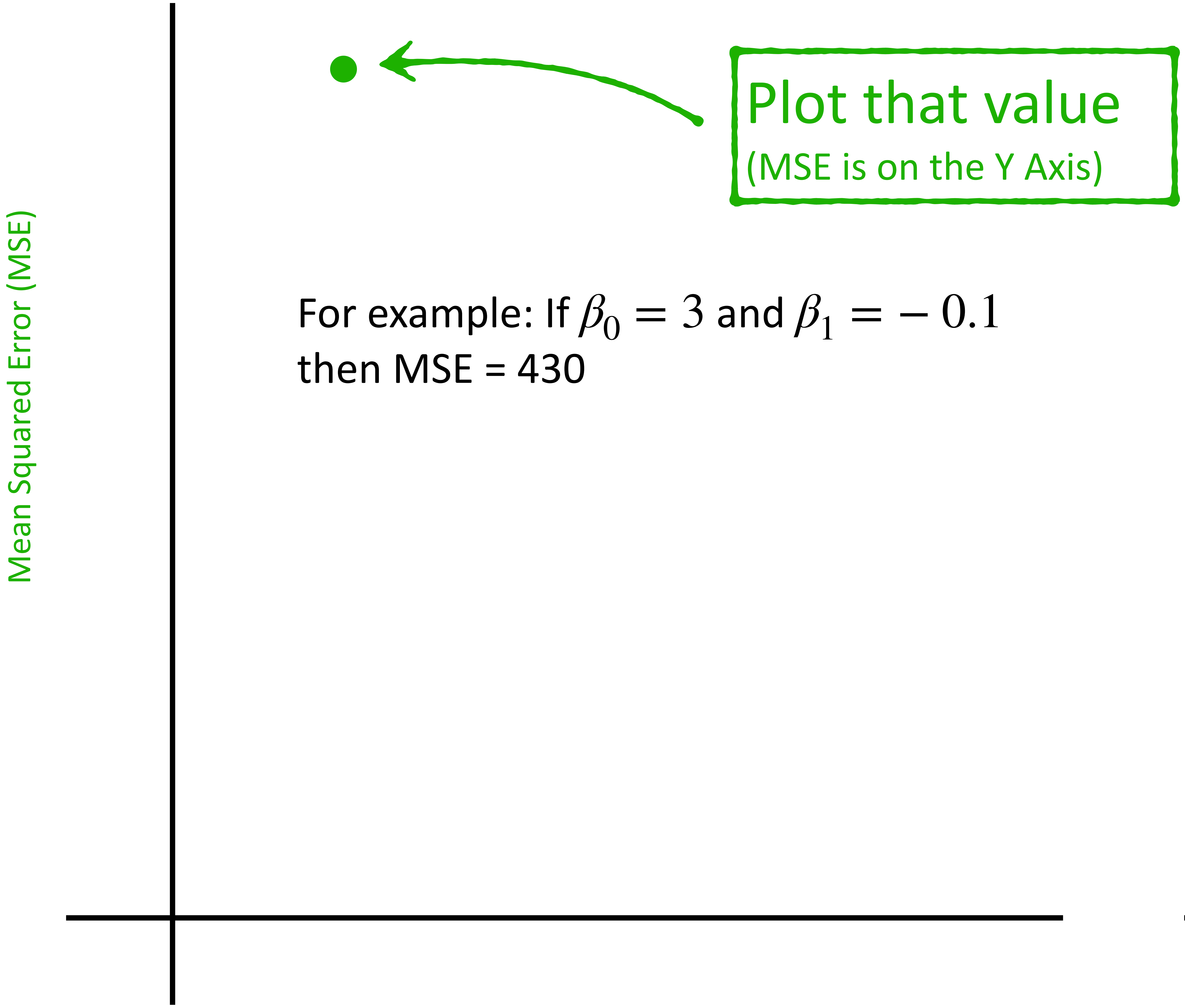
The Problem Statement:

Simple Linear Regression: Find the values of β_0 and β_1 such that the Mean Squared Error (MSE) is minimized.

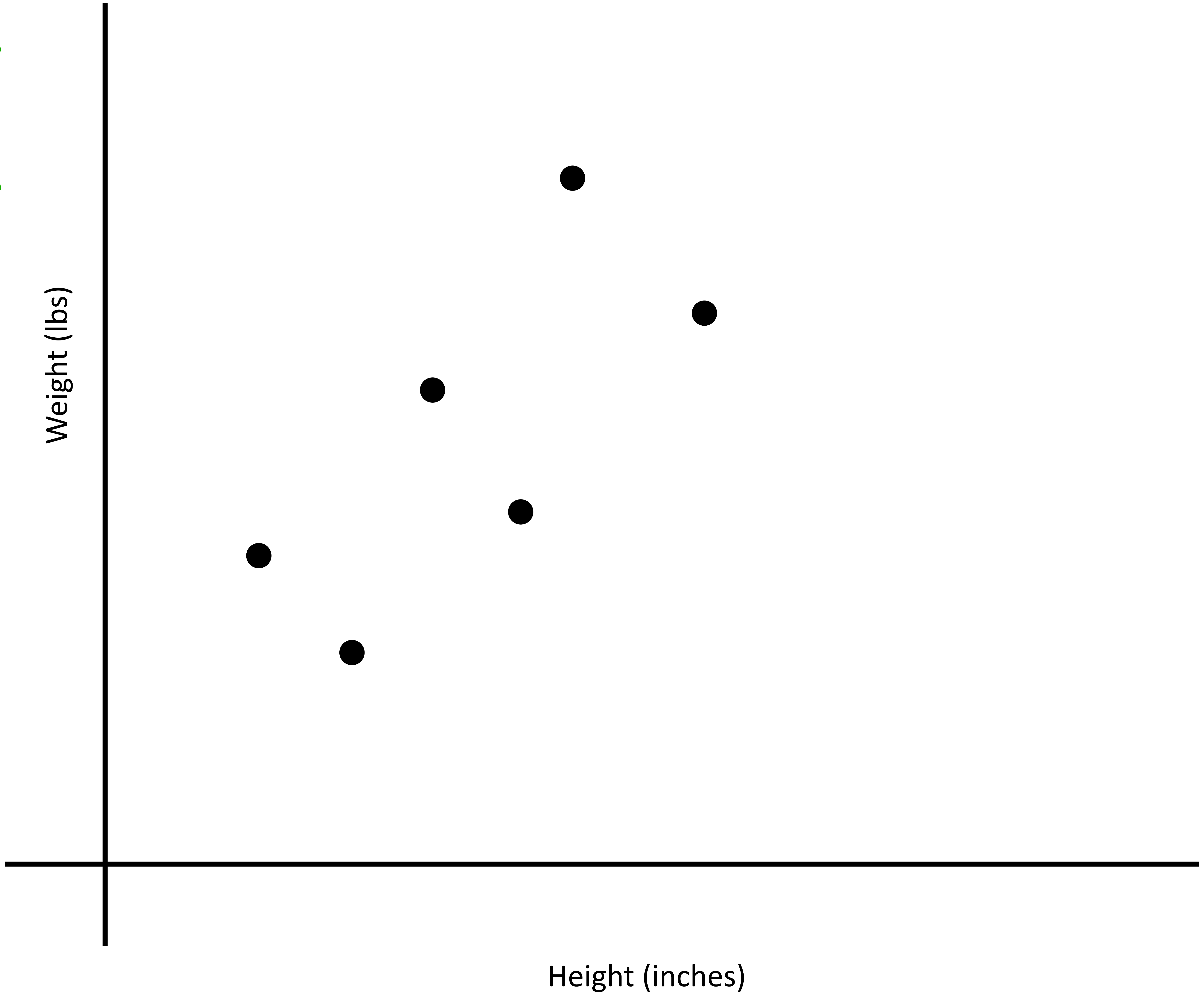


i	Height (in)	Weight (lbs)
1	62	138
2	55	178
3	44	123
4	75	200
5	65	229
6	50	102

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

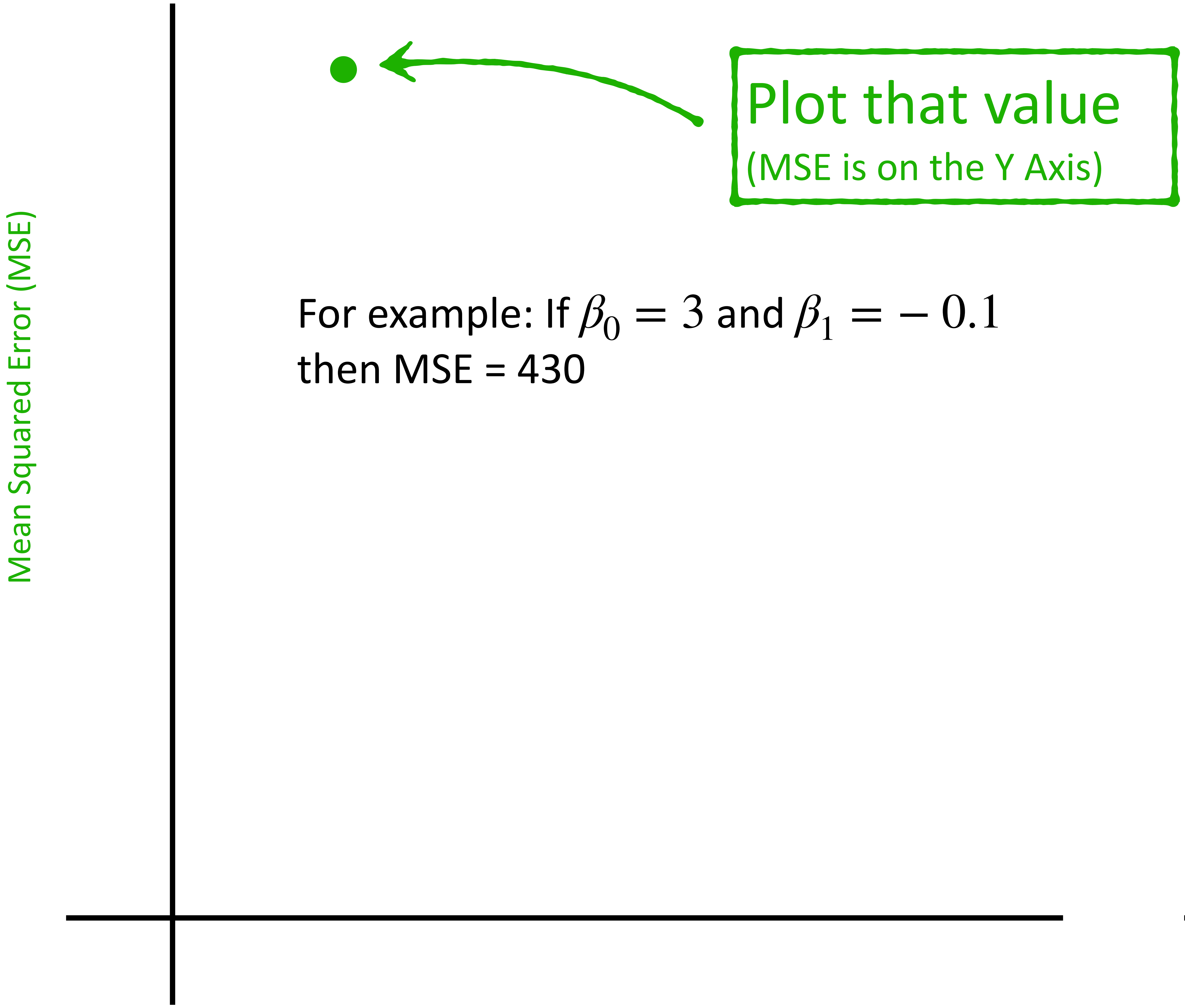


For example: If $\beta_0 = 3$ and $\beta_1 = -0.1$ then MSE = 430

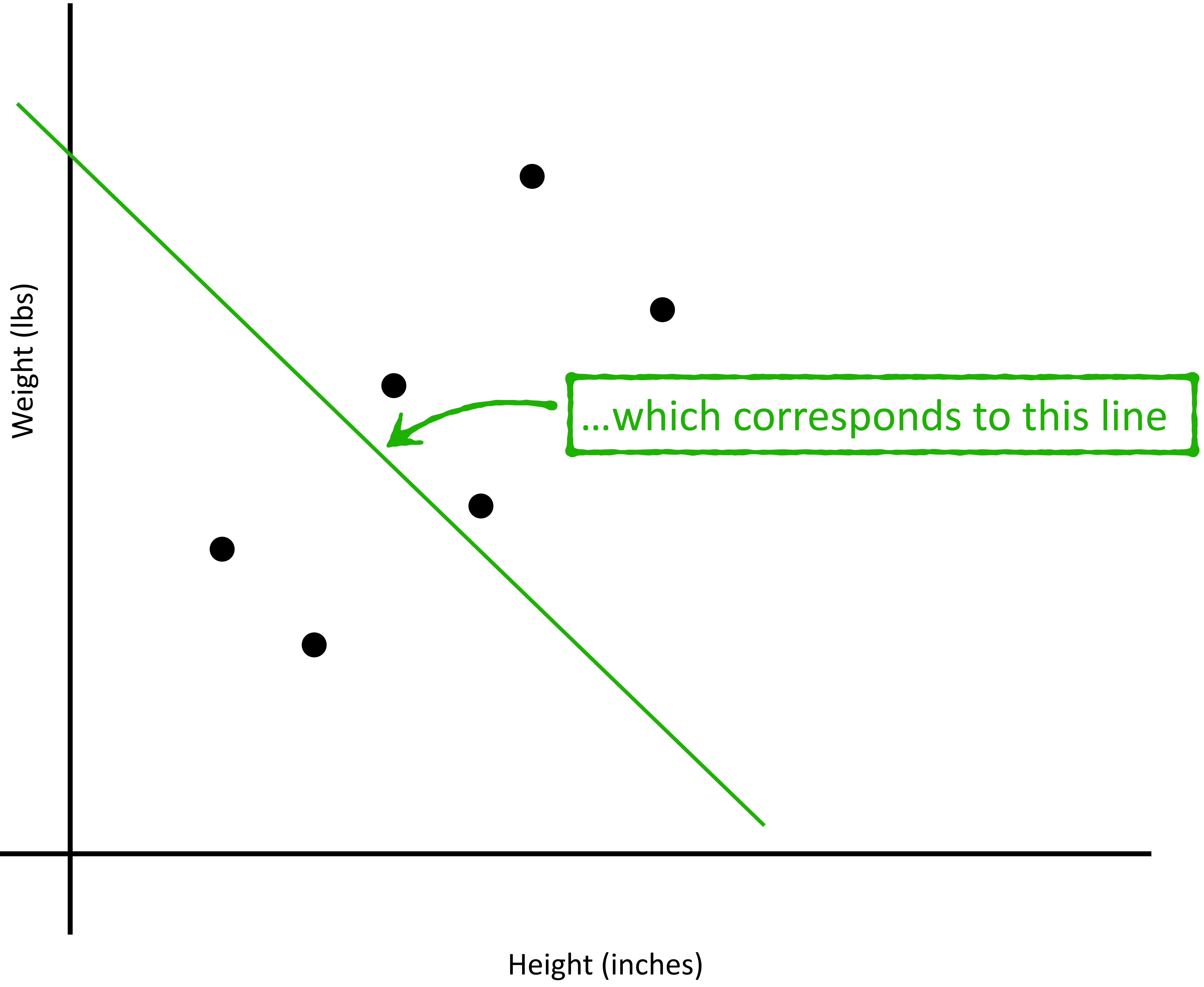


Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

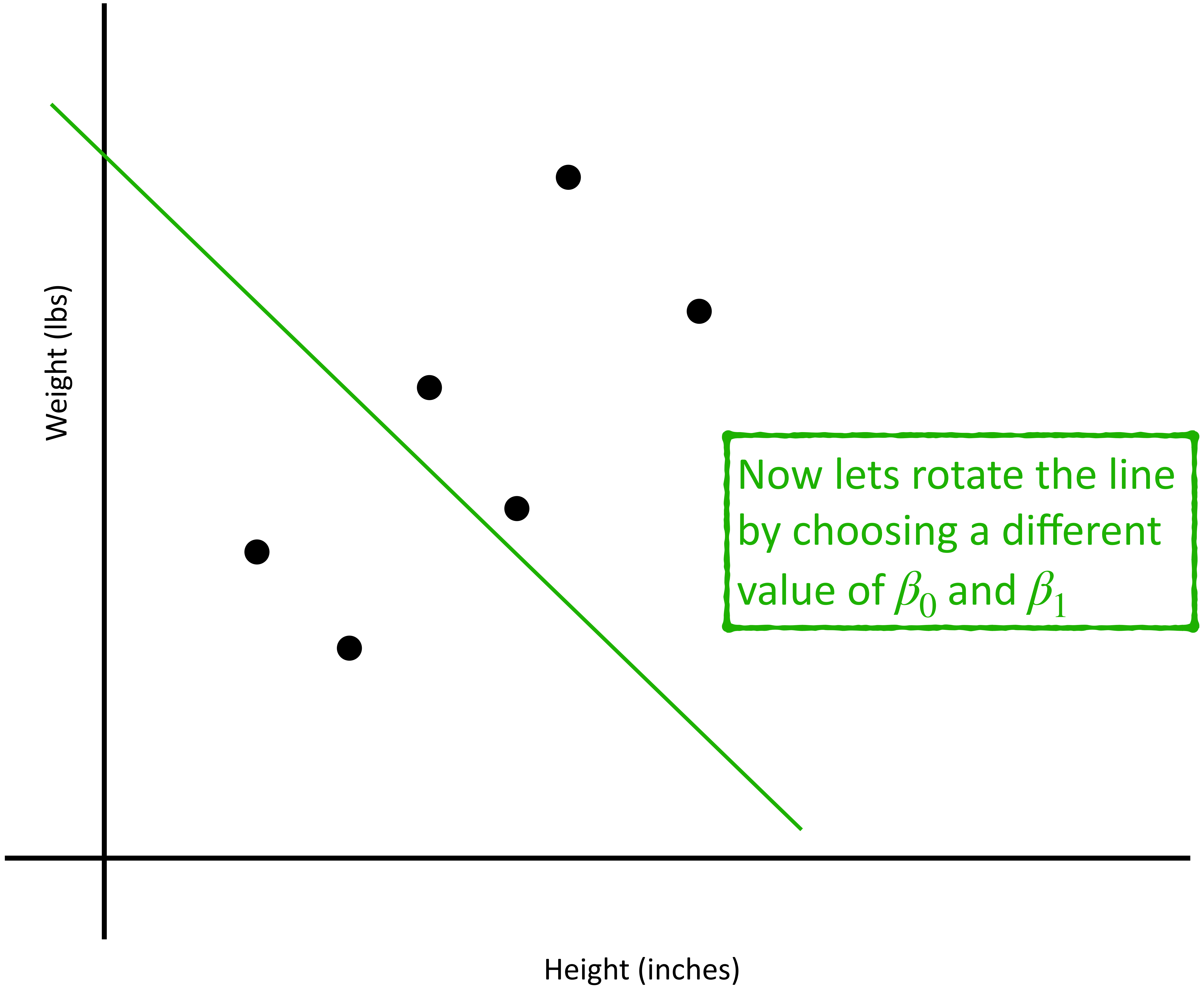


For example: If $\beta_0 = 3$ and $\beta_1 = -0.1$
then MSE = 430



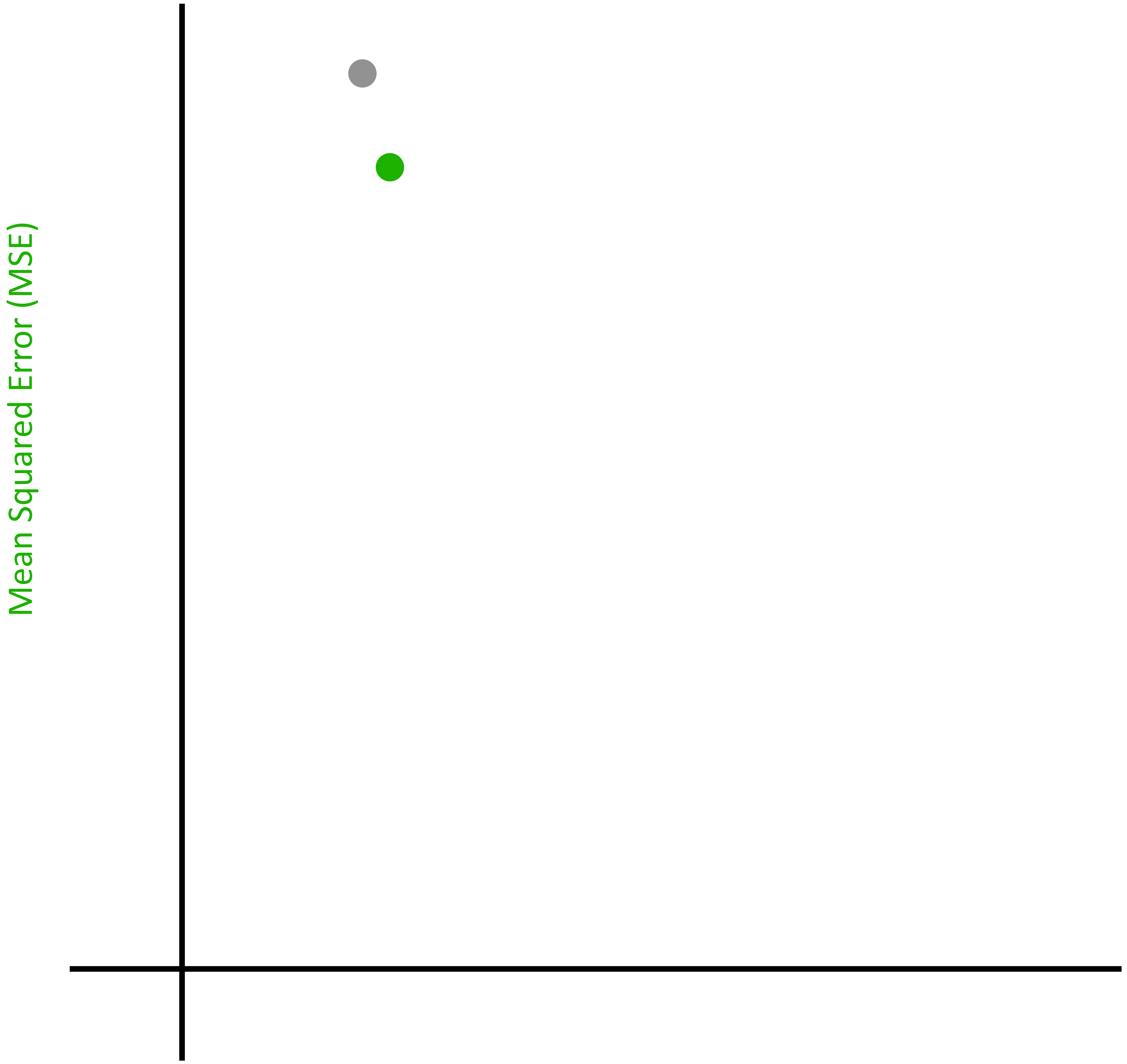
Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

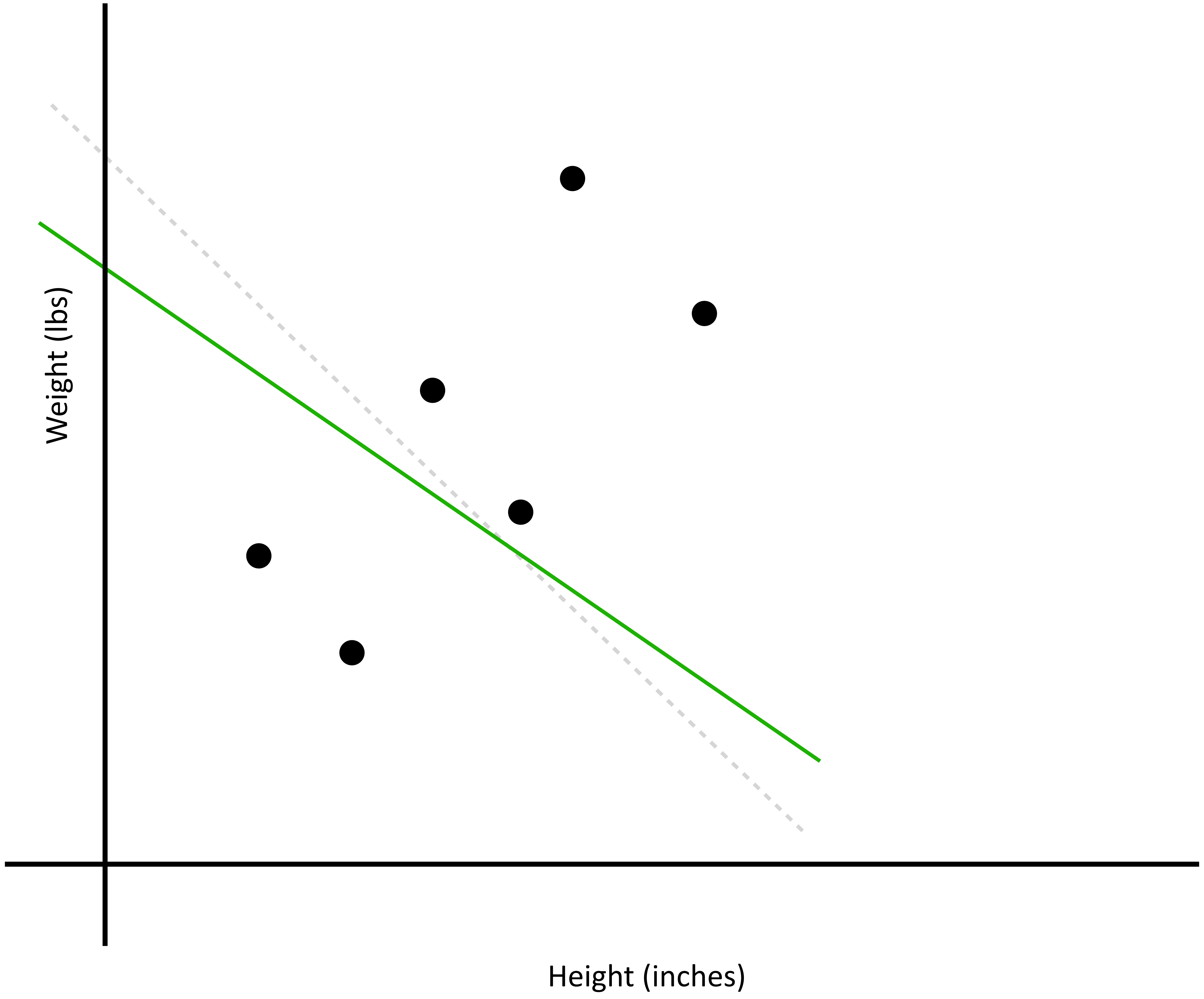
Simple Linear Regression



Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

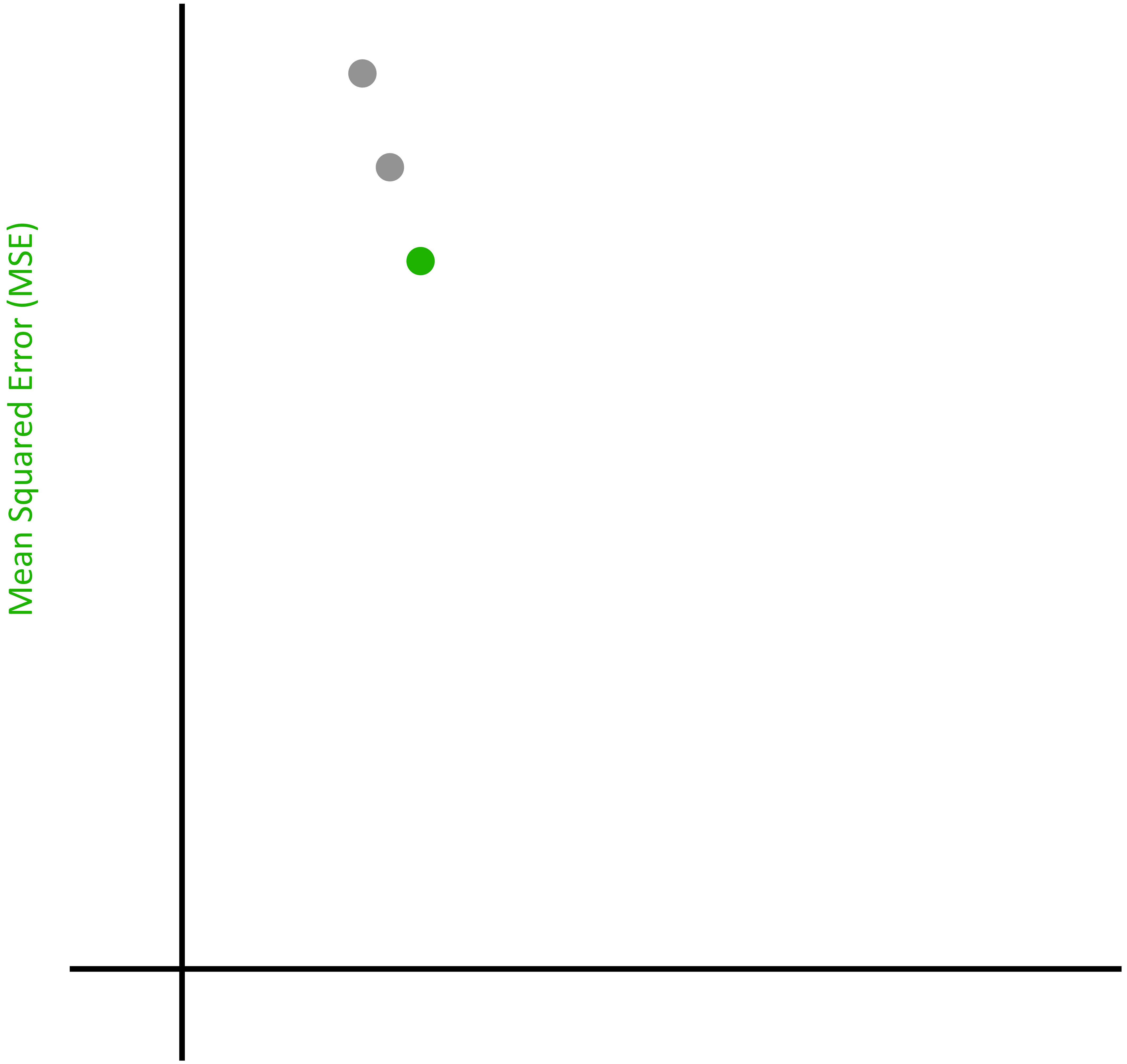
Simple Linear Regression

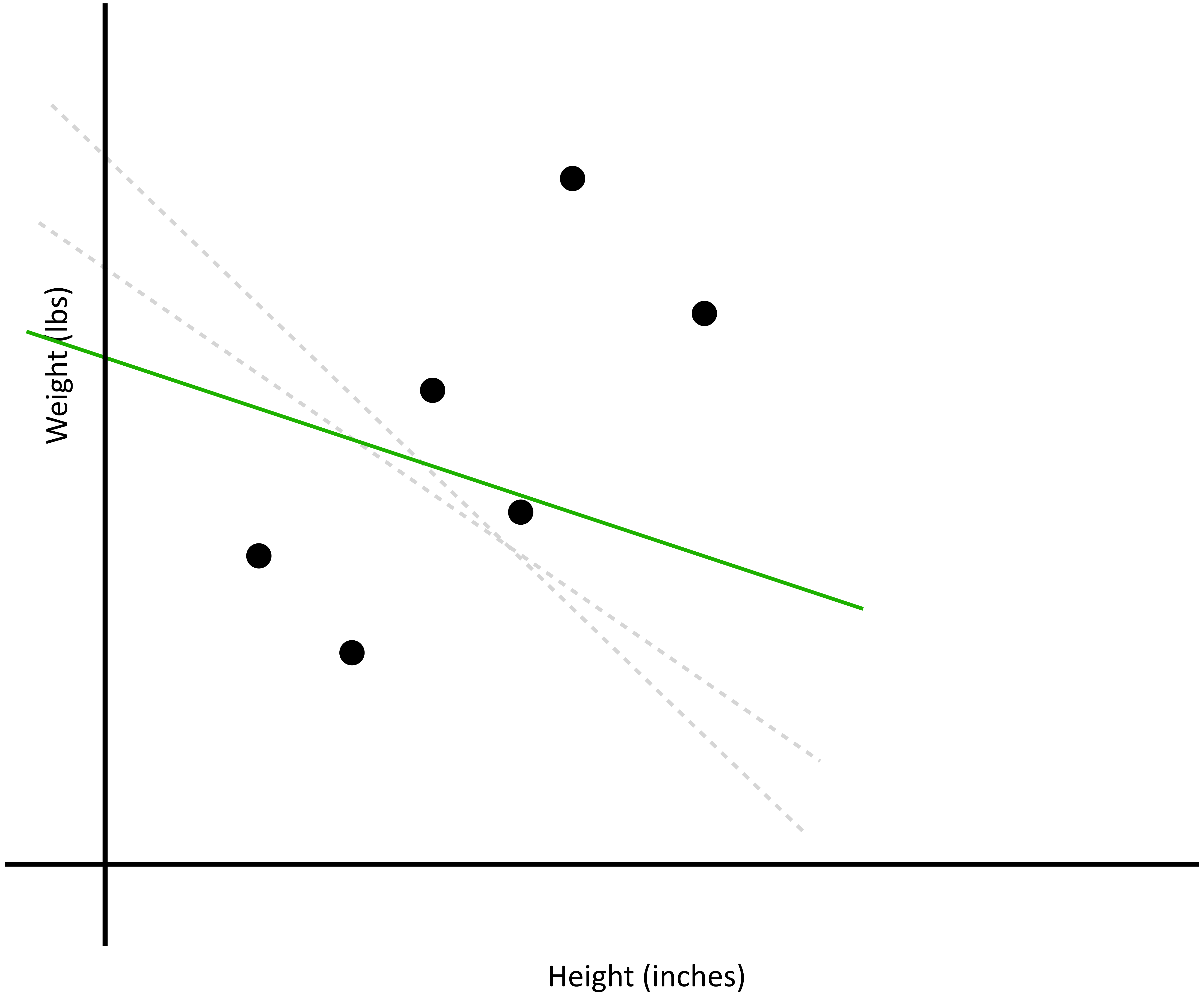




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

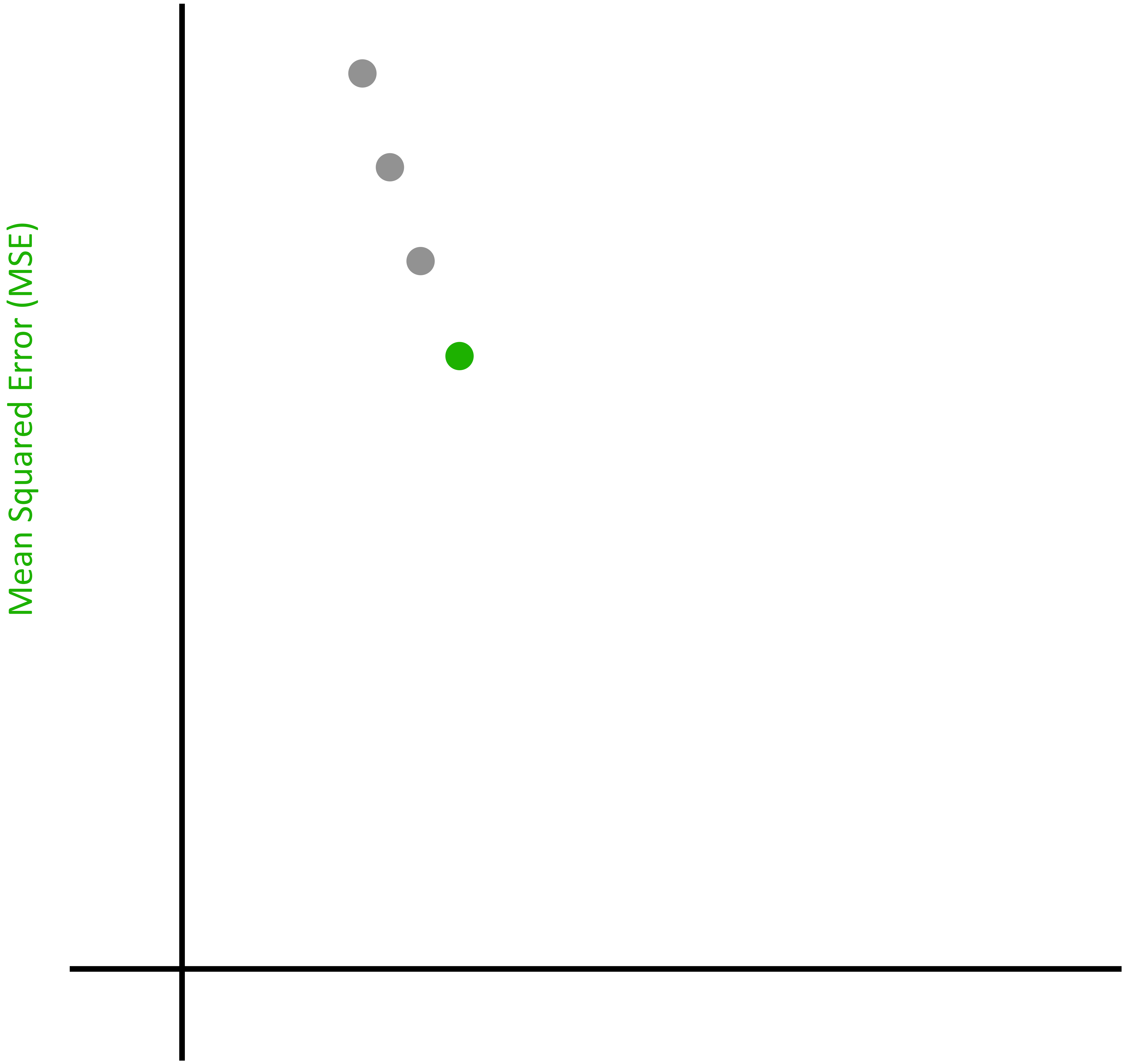
Simple Linear Regression

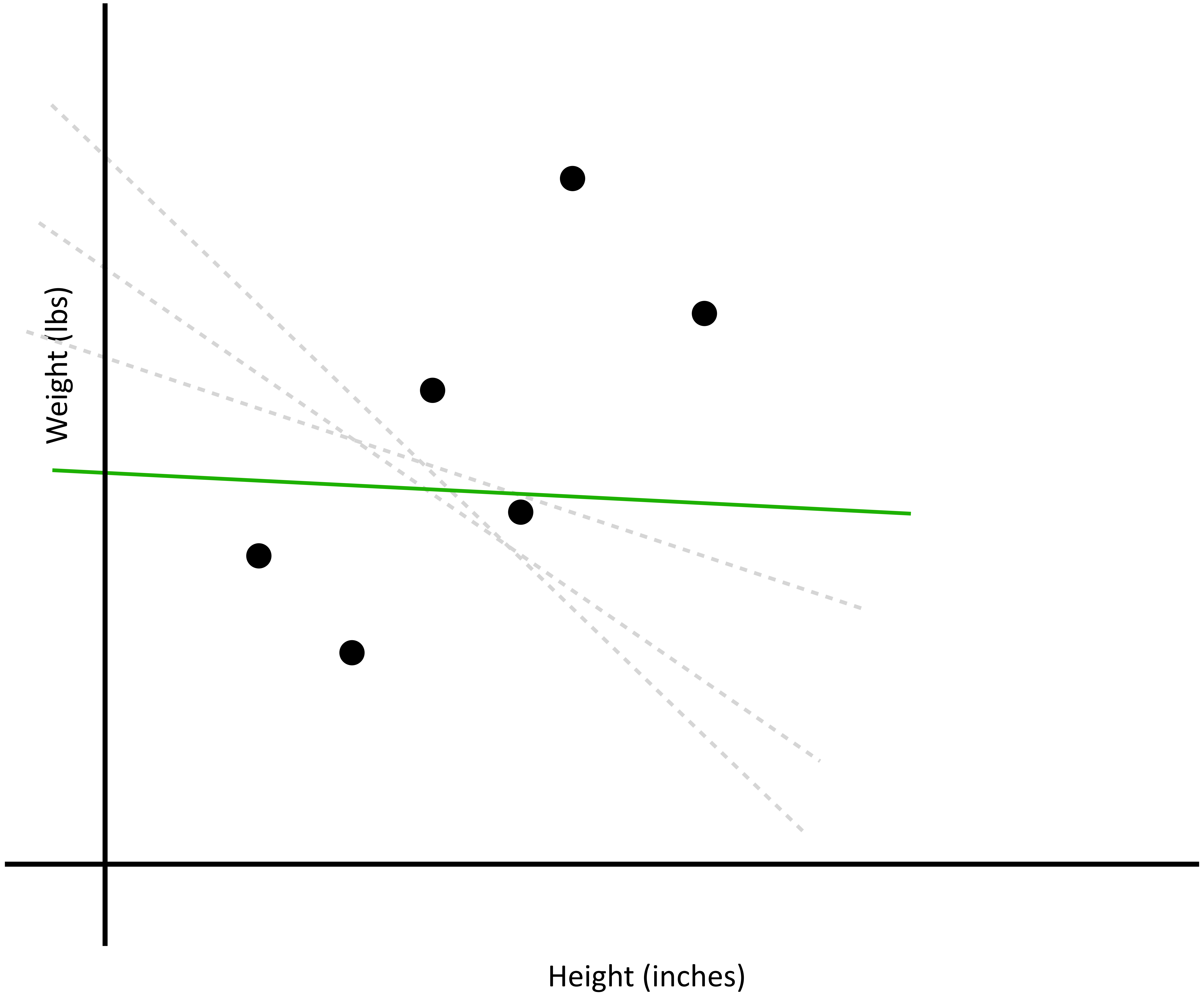




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

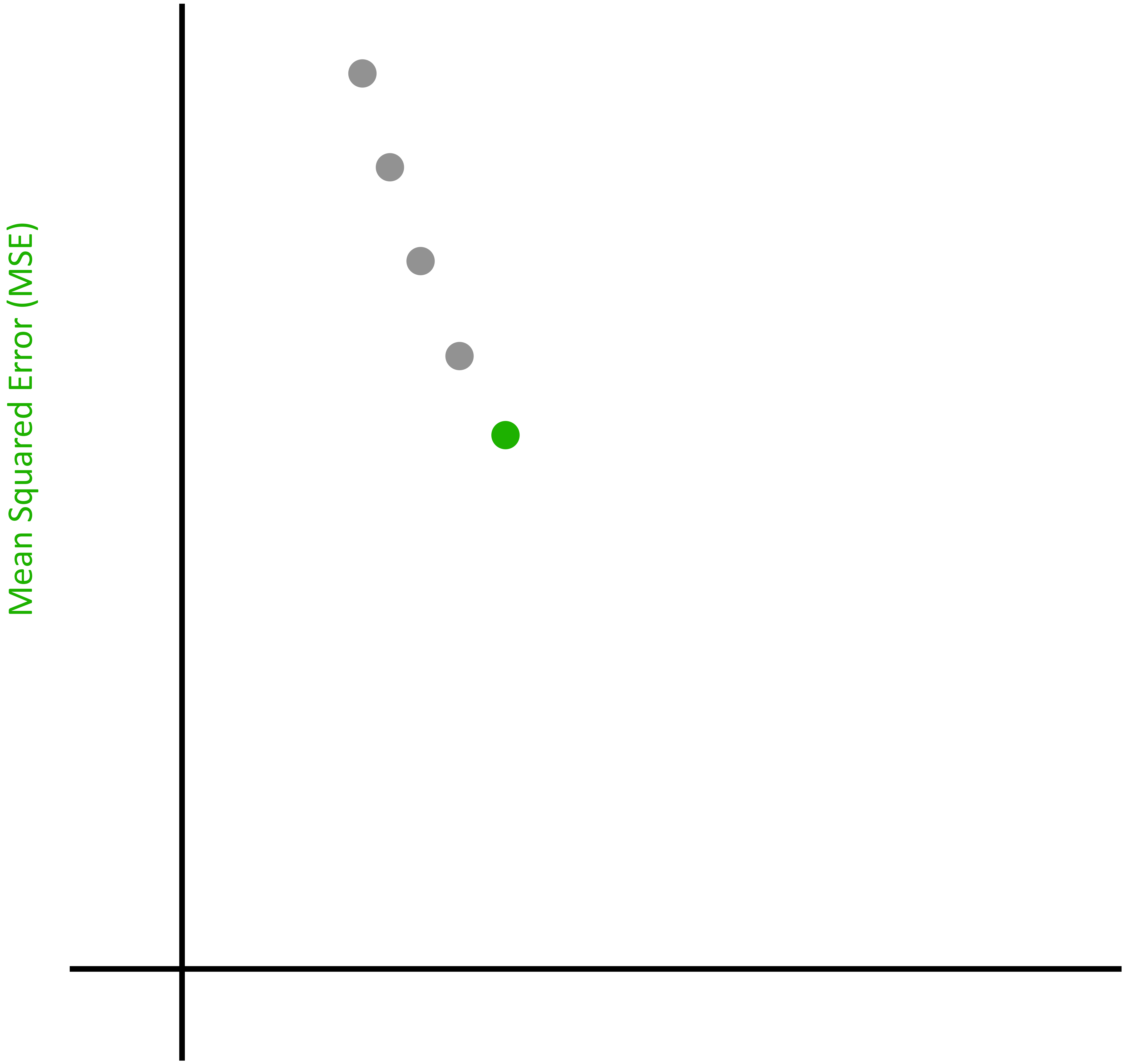
Simple Linear Regression

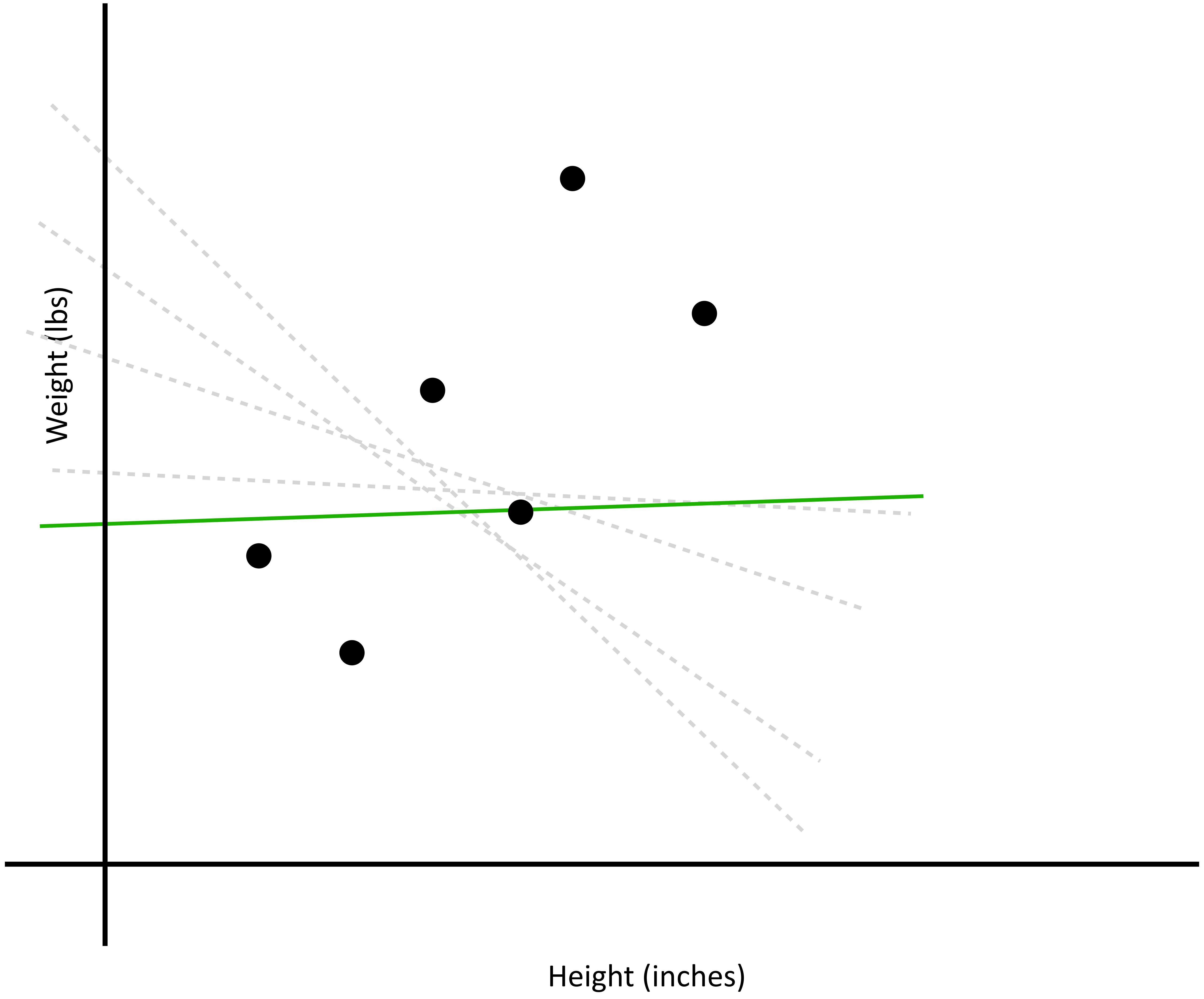




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

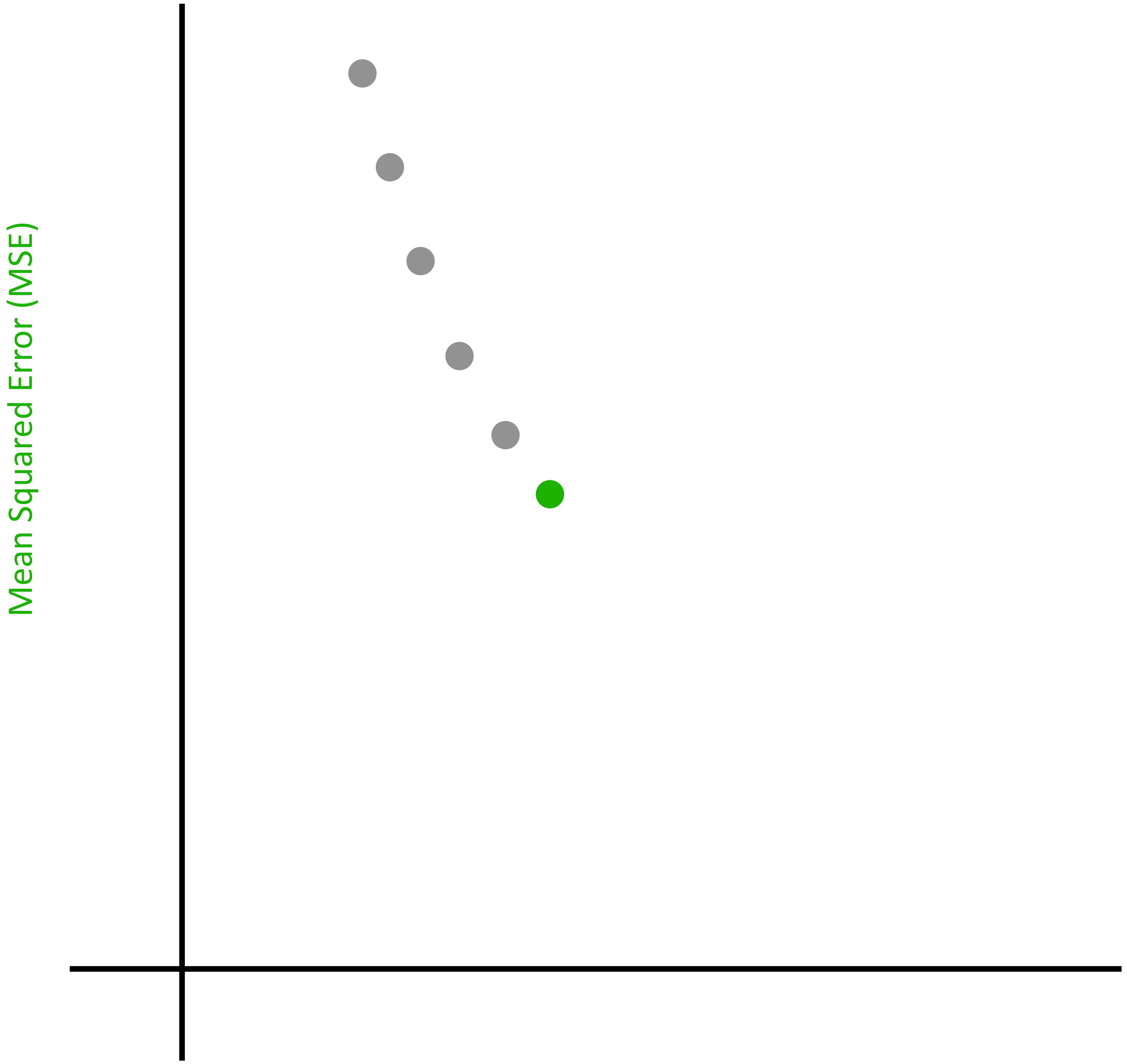
Simple Linear Regression

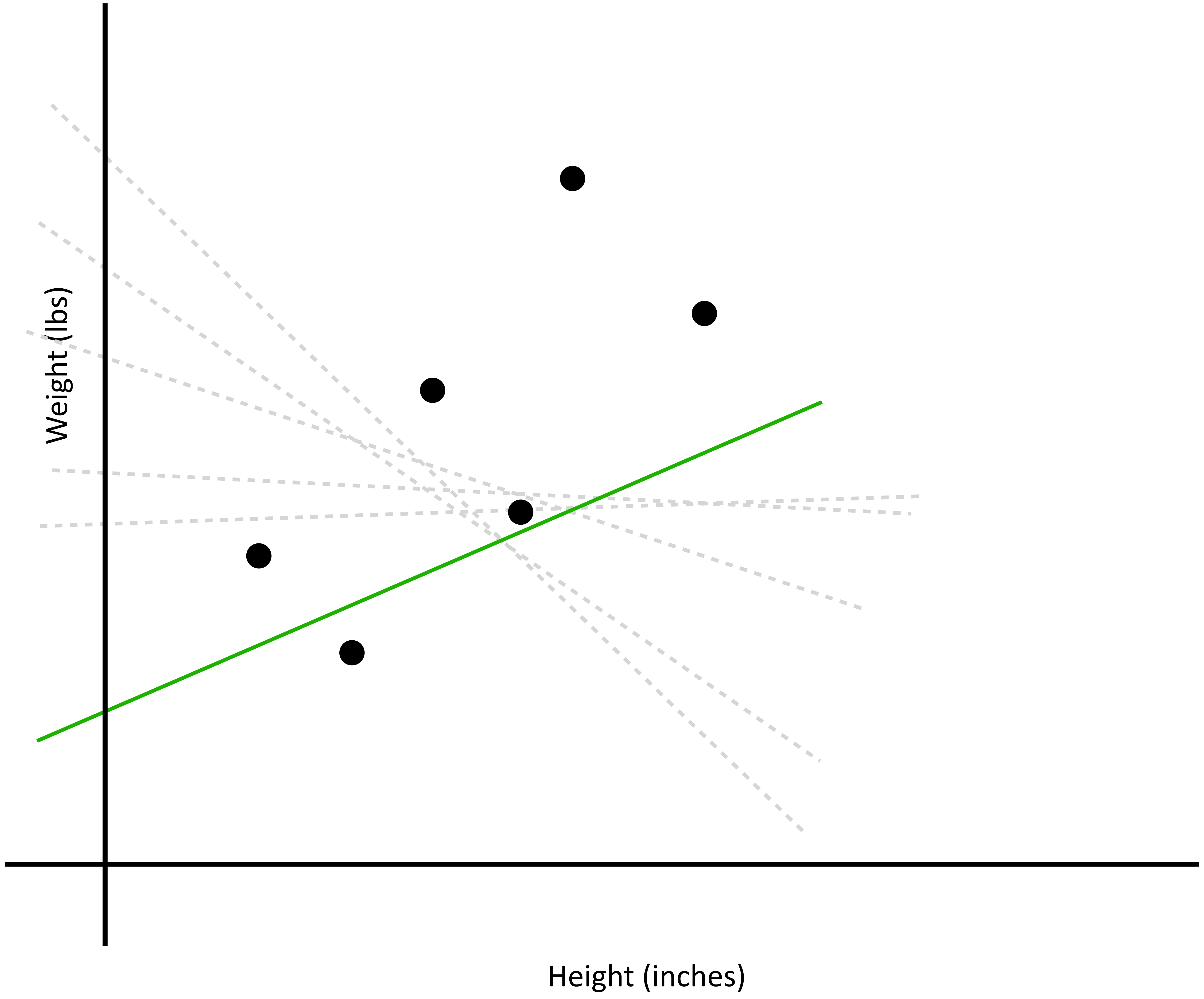




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

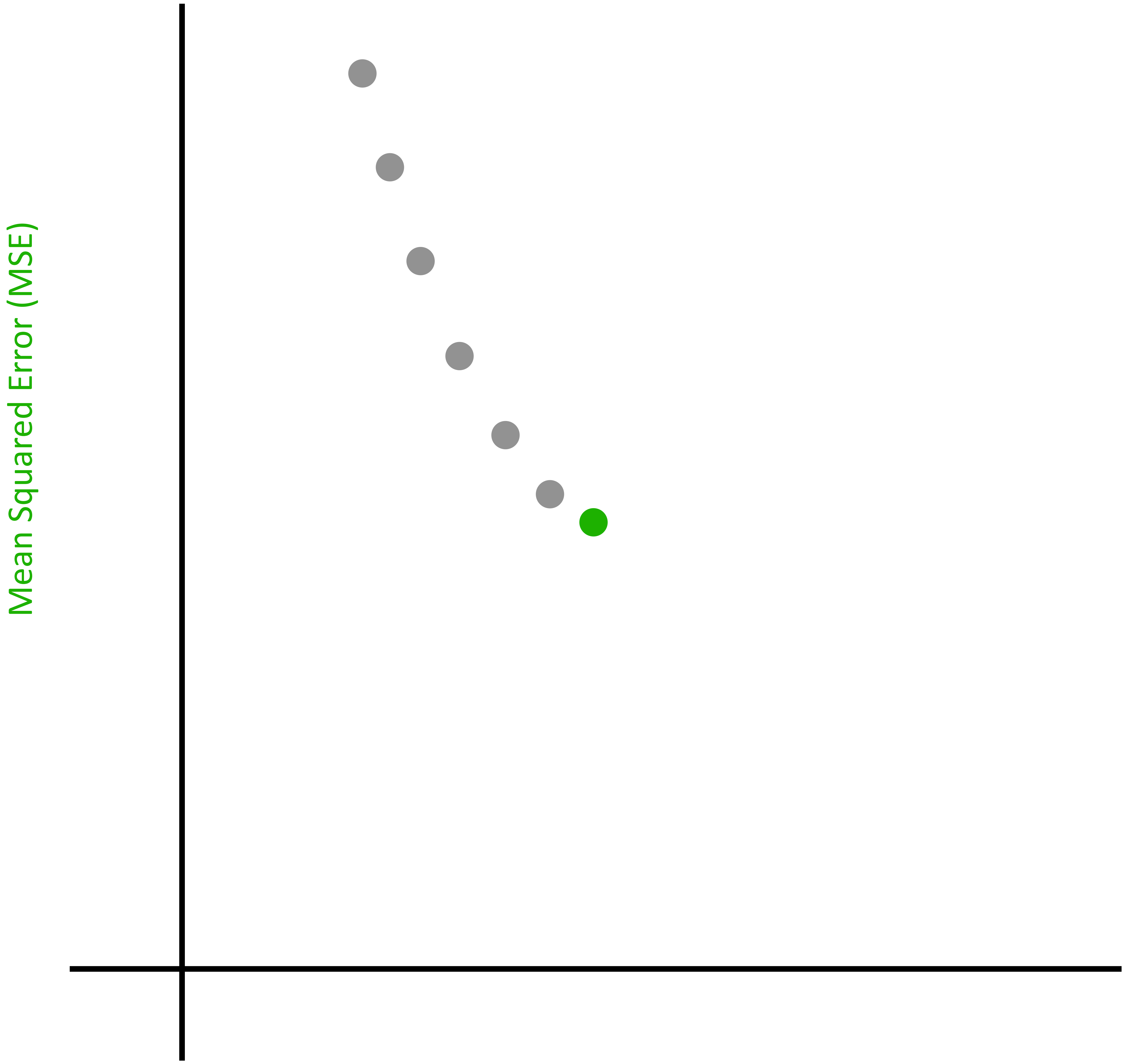
Simple Linear Regression

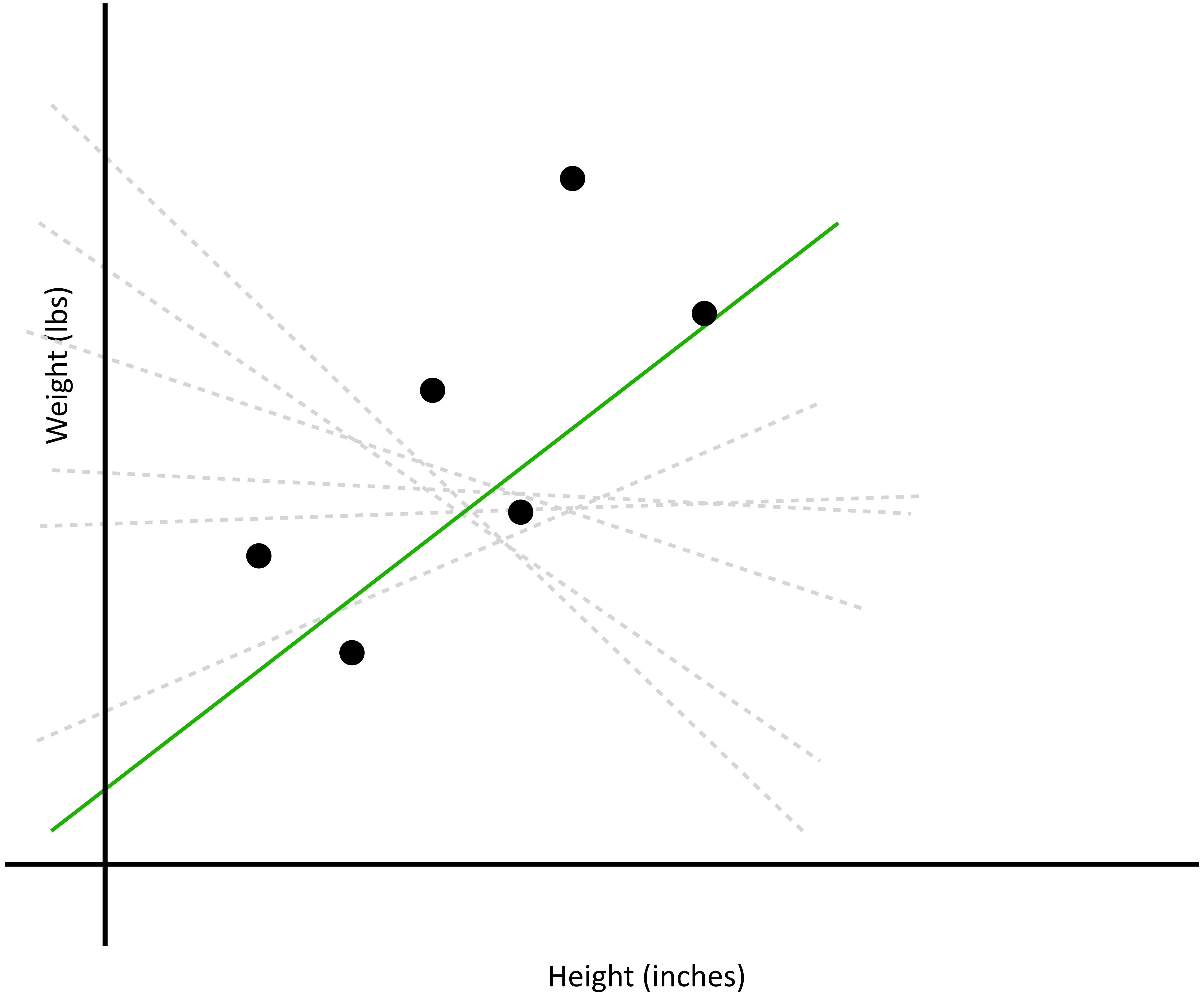




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

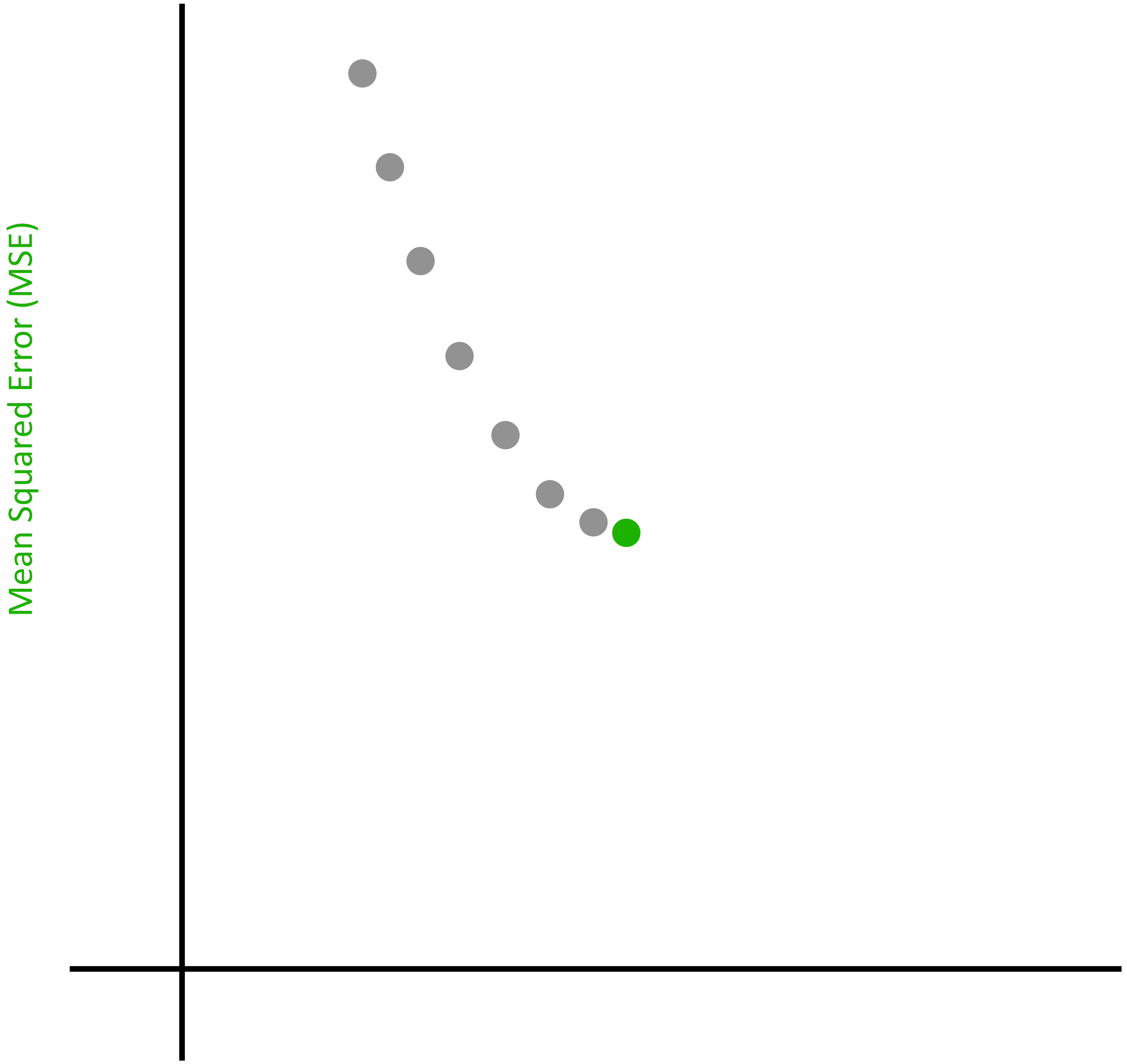
Simple Linear Regression

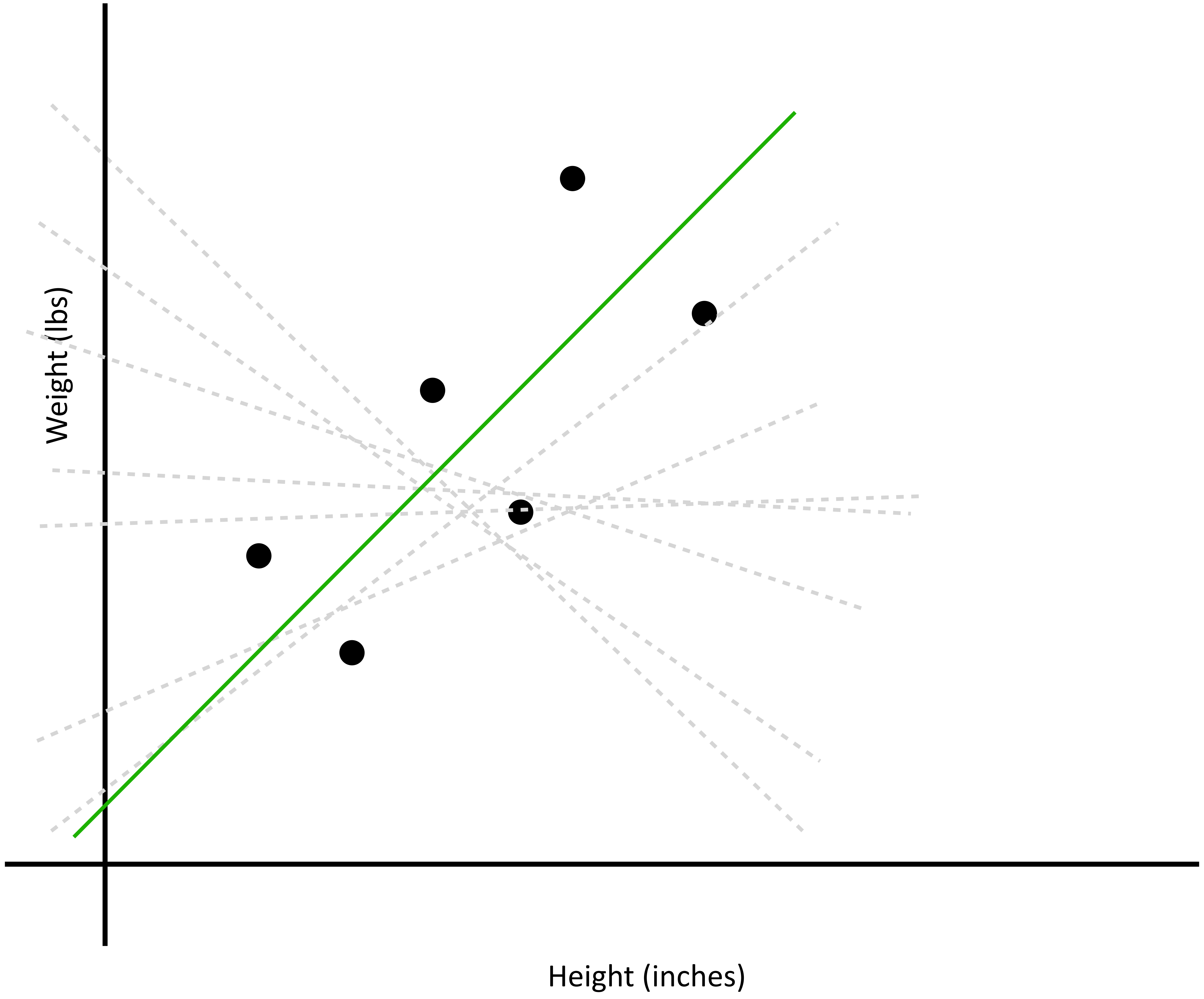




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

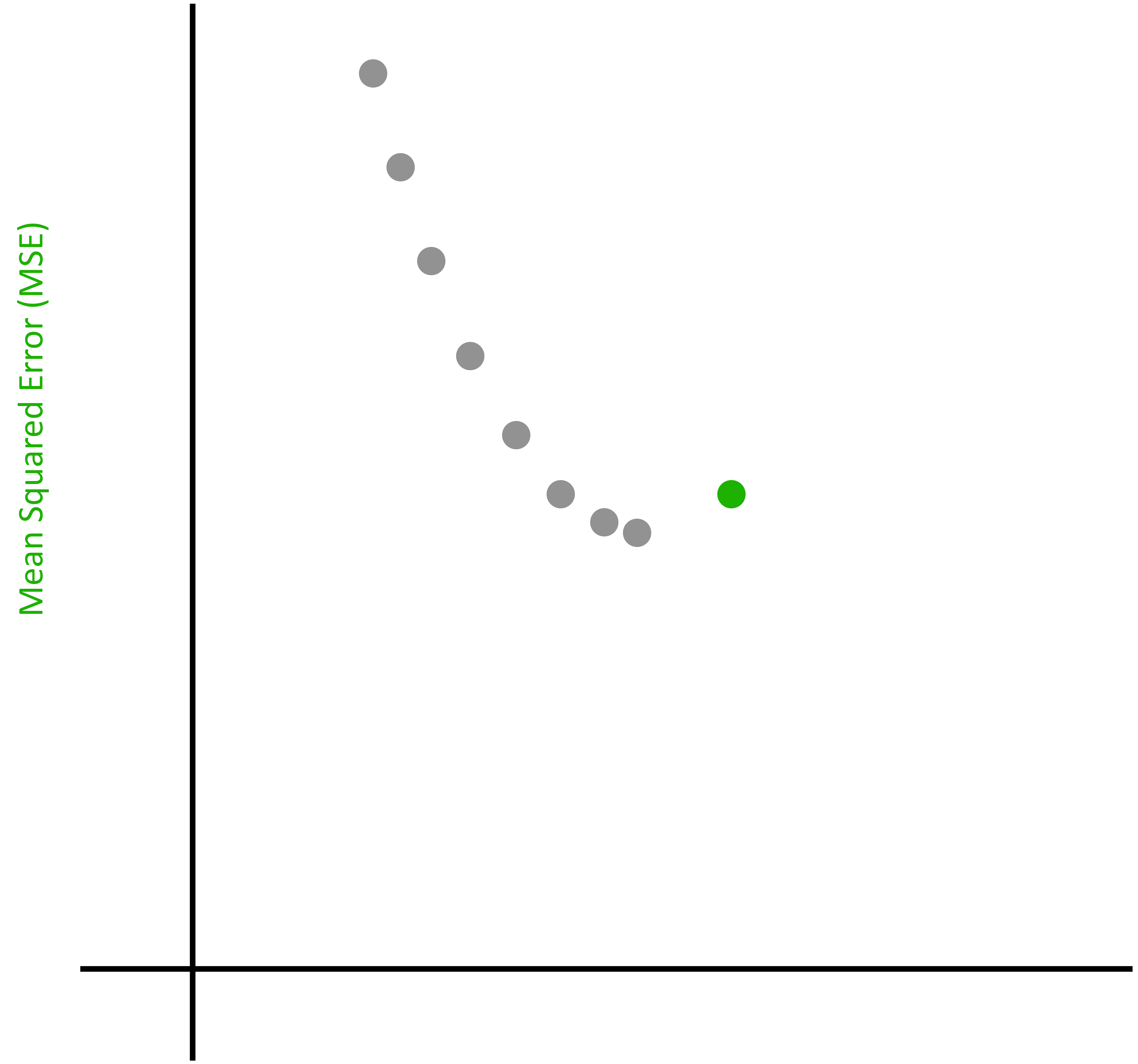
Simple Linear Regression

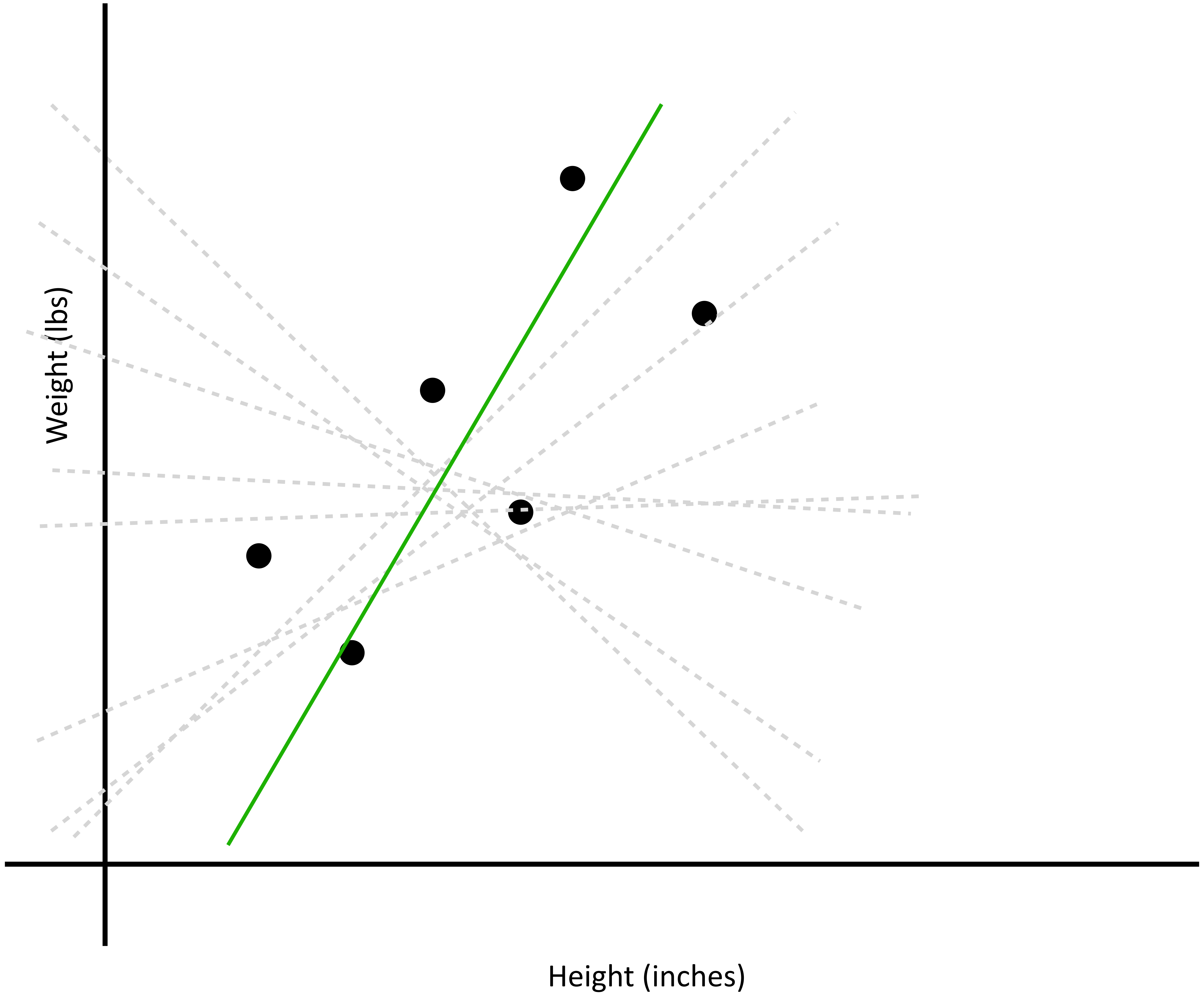




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

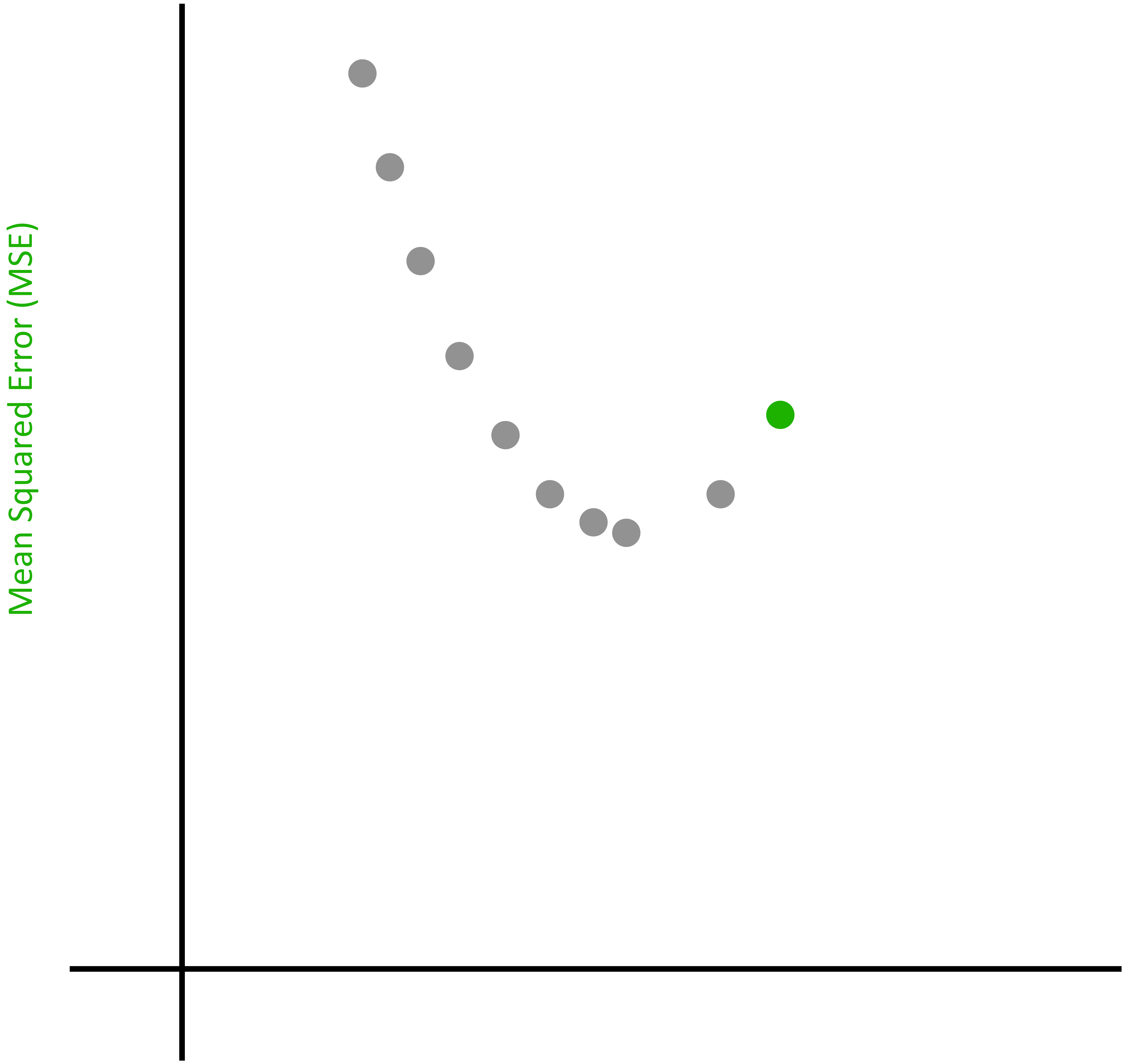
Simple Linear Regression

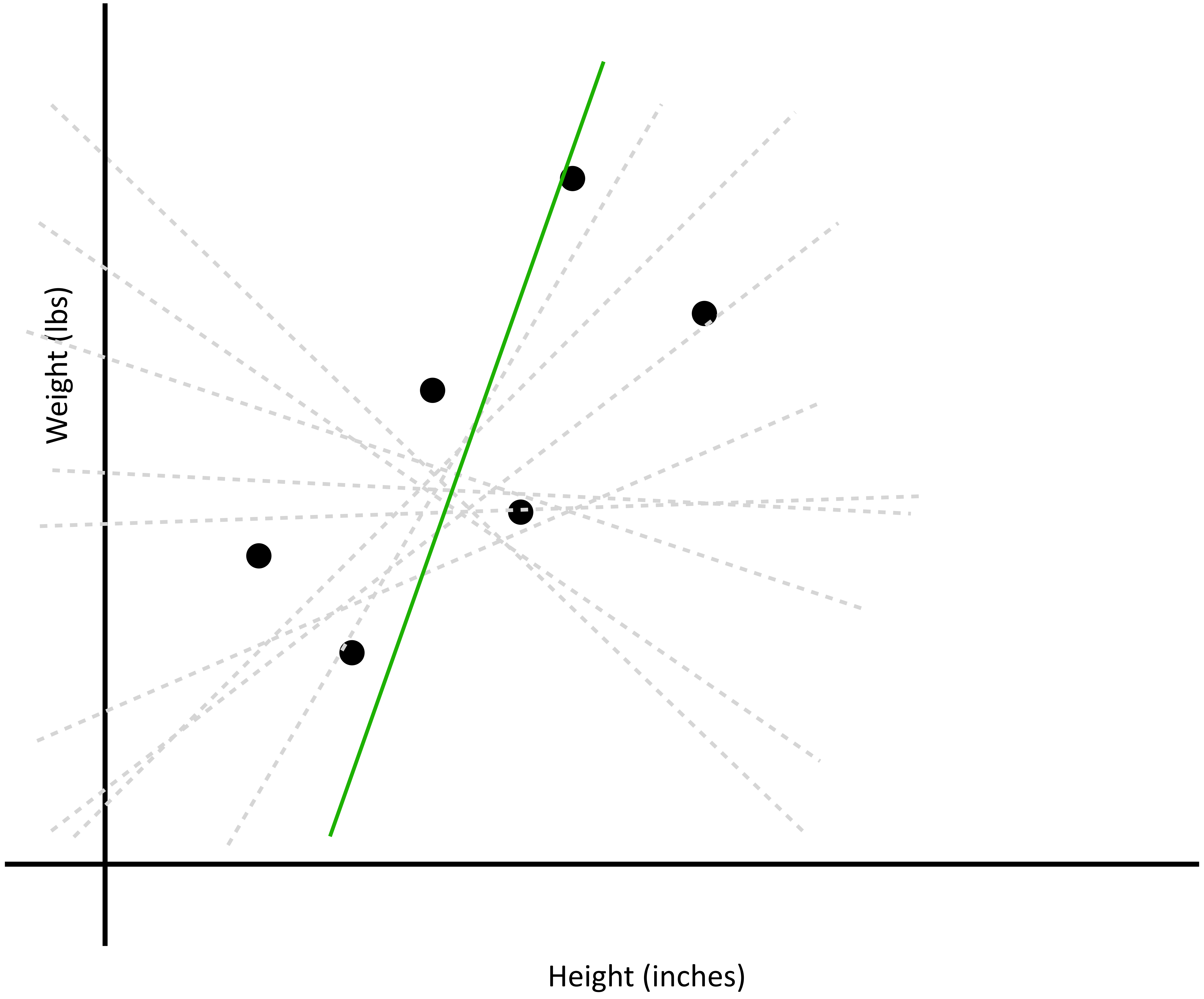




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

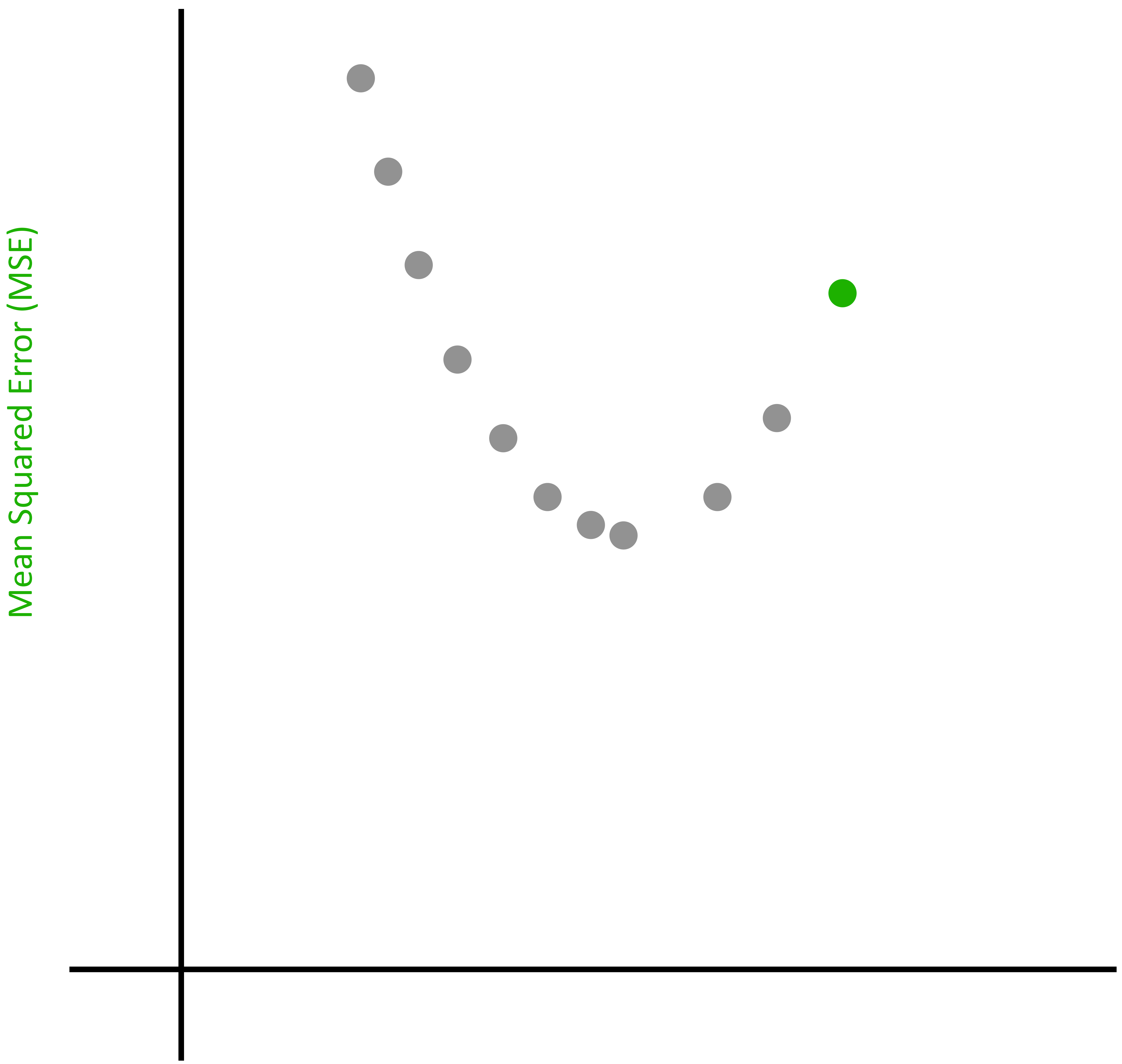
Simple Linear Regression

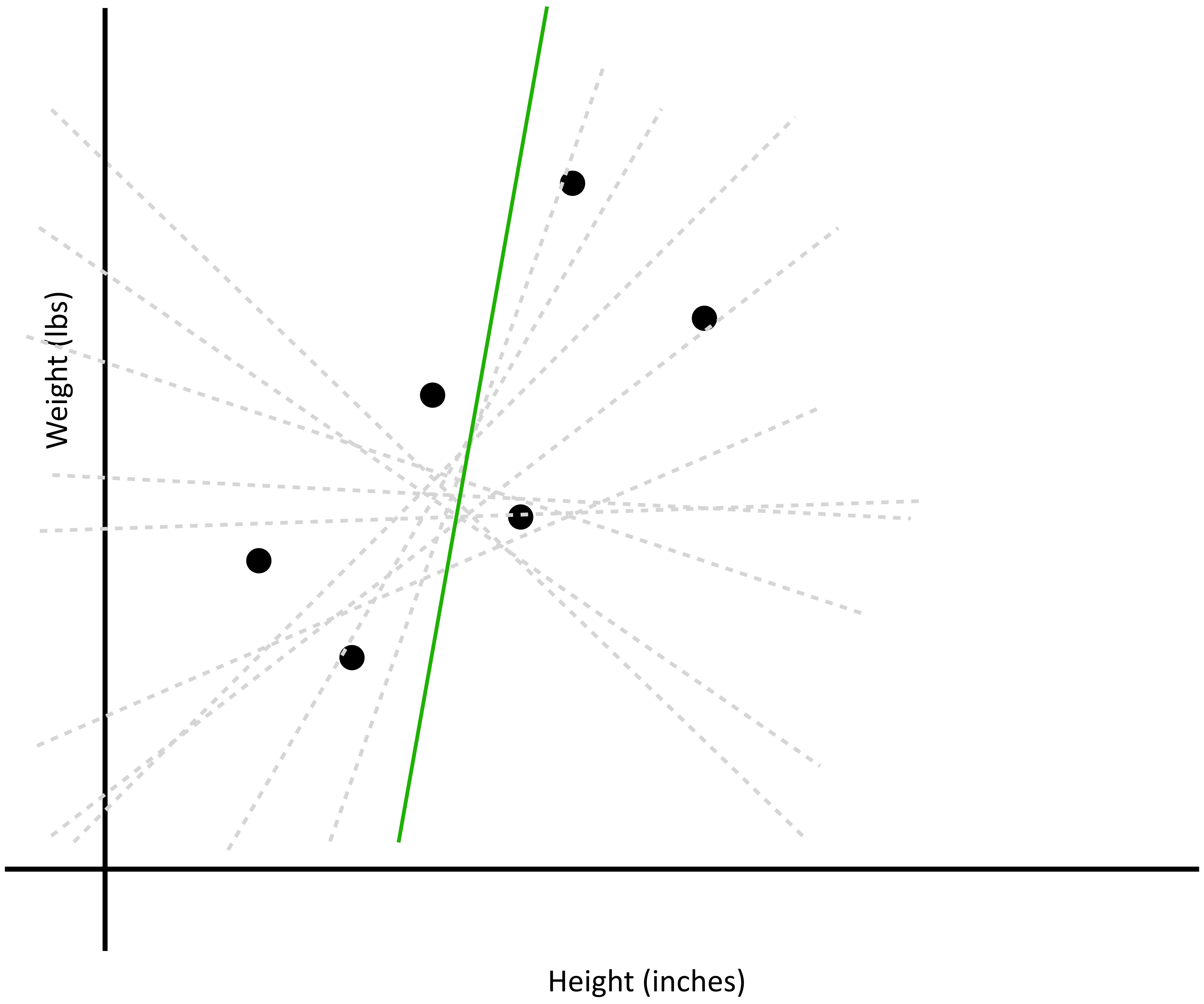




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

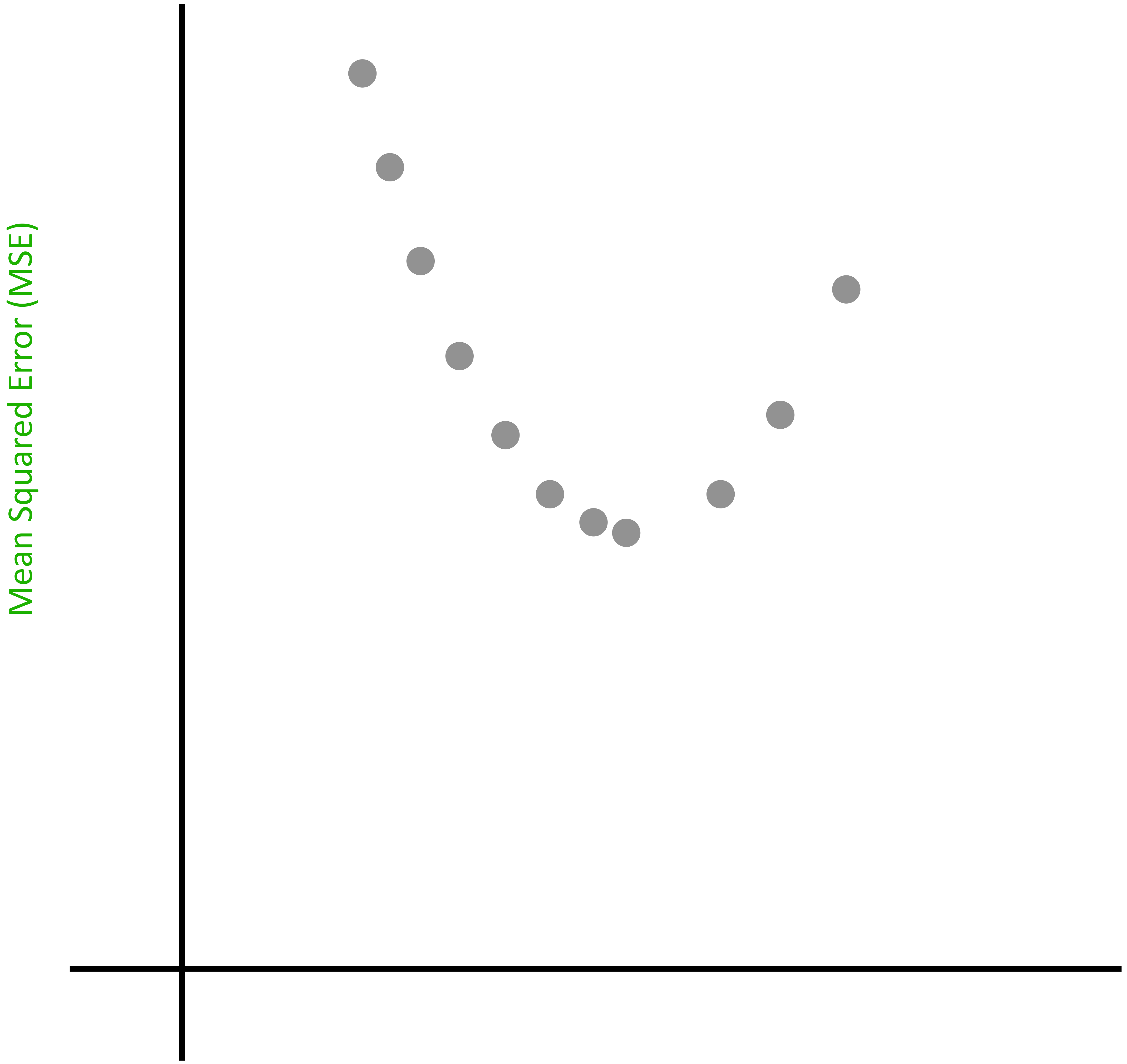
Simple Linear Regression

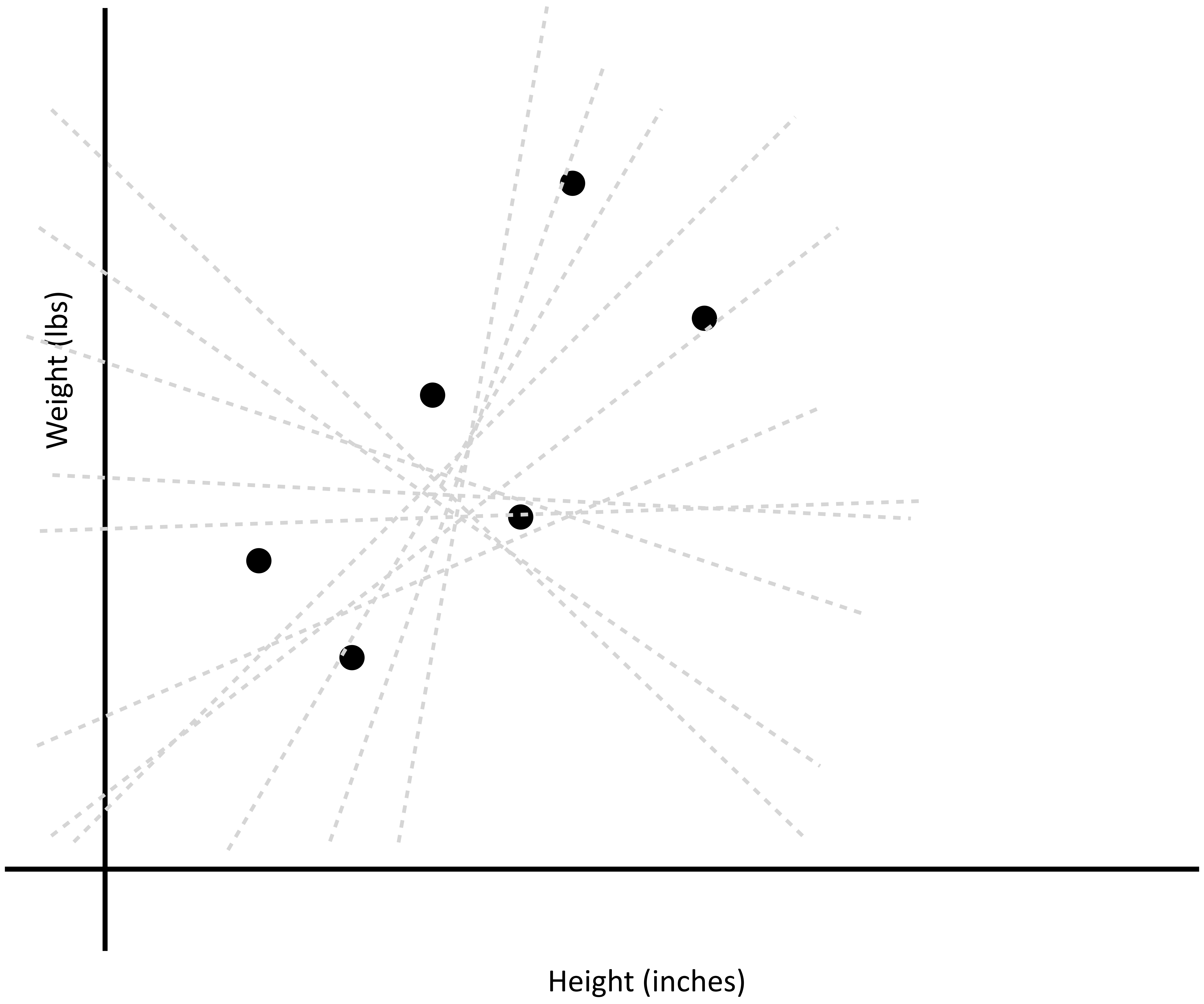




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

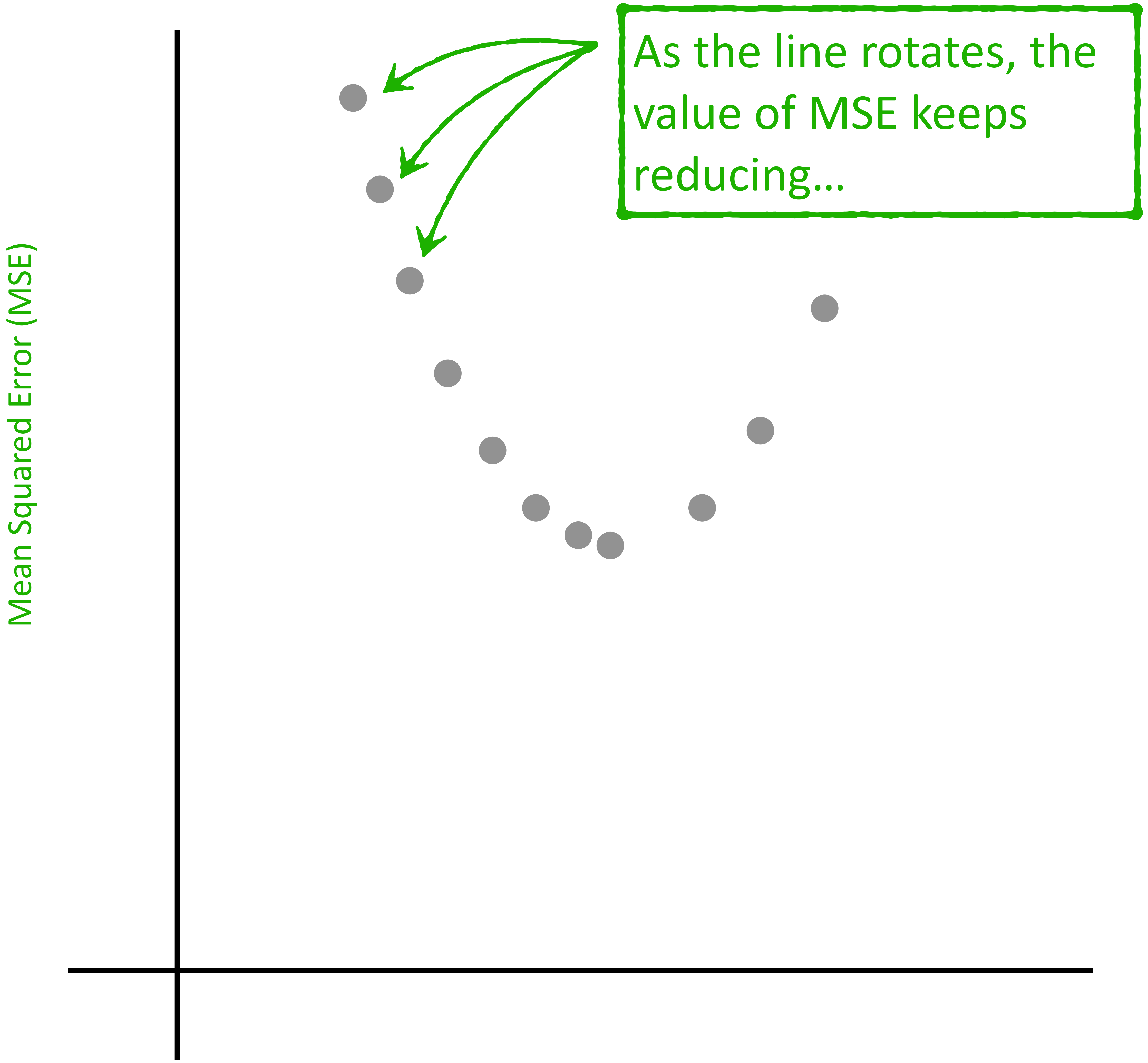
Simple Linear Regression

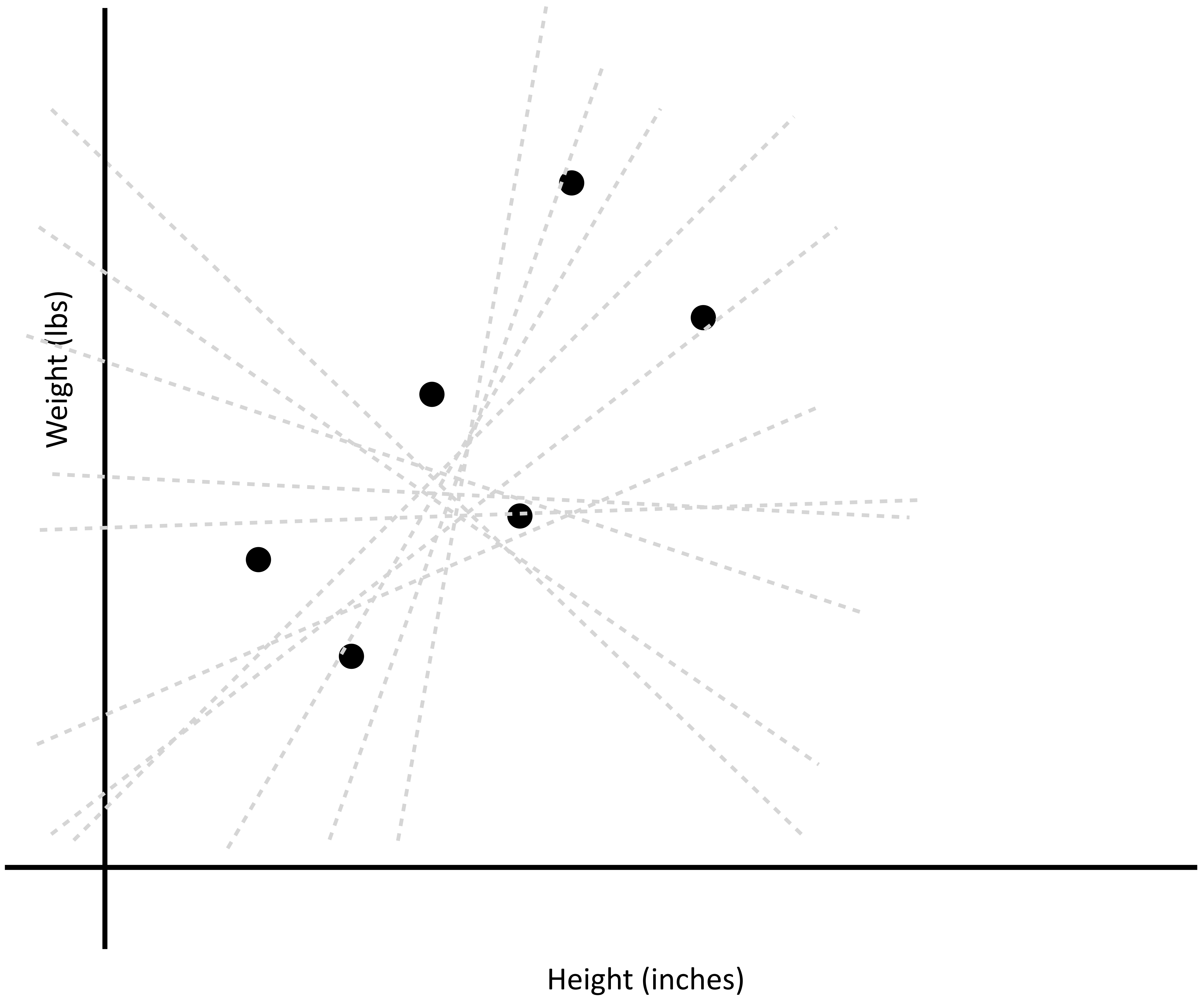




Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

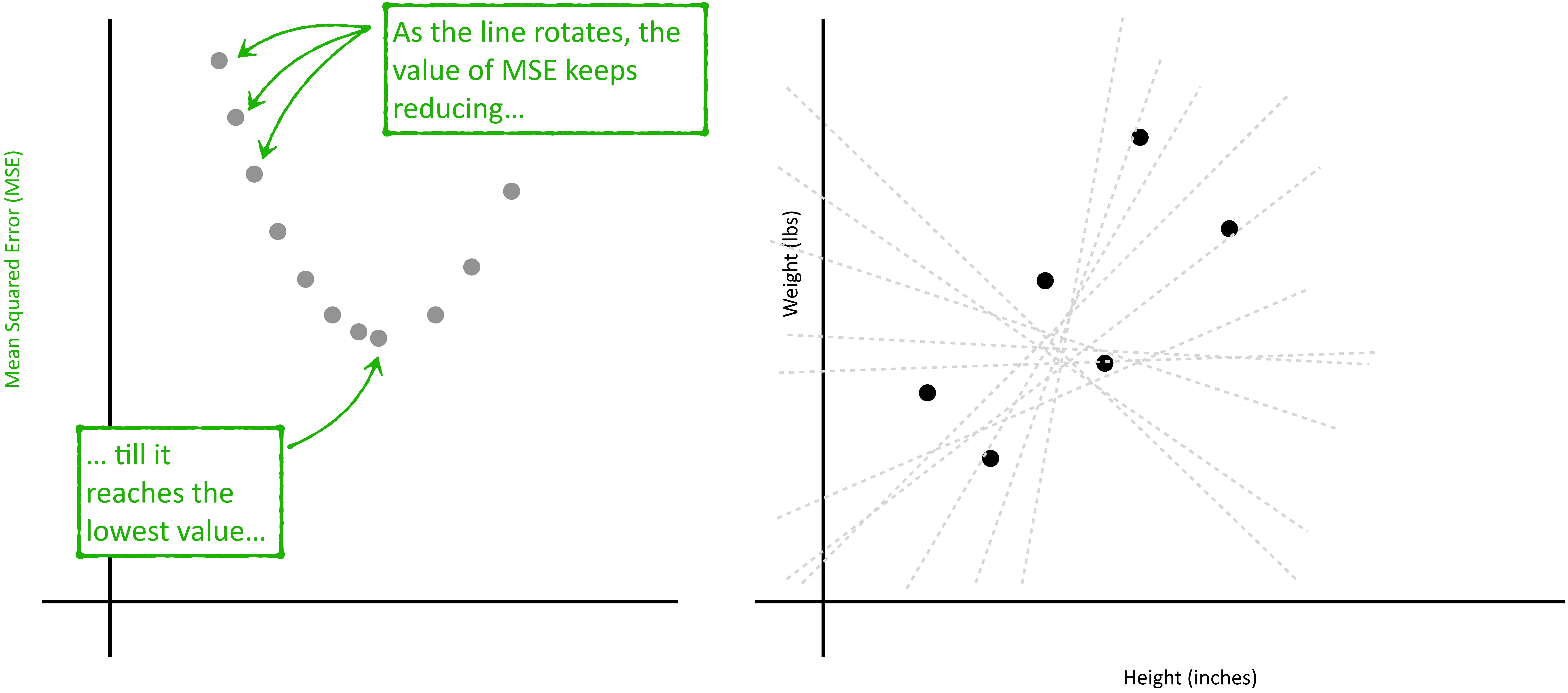
Simple Linear Regression





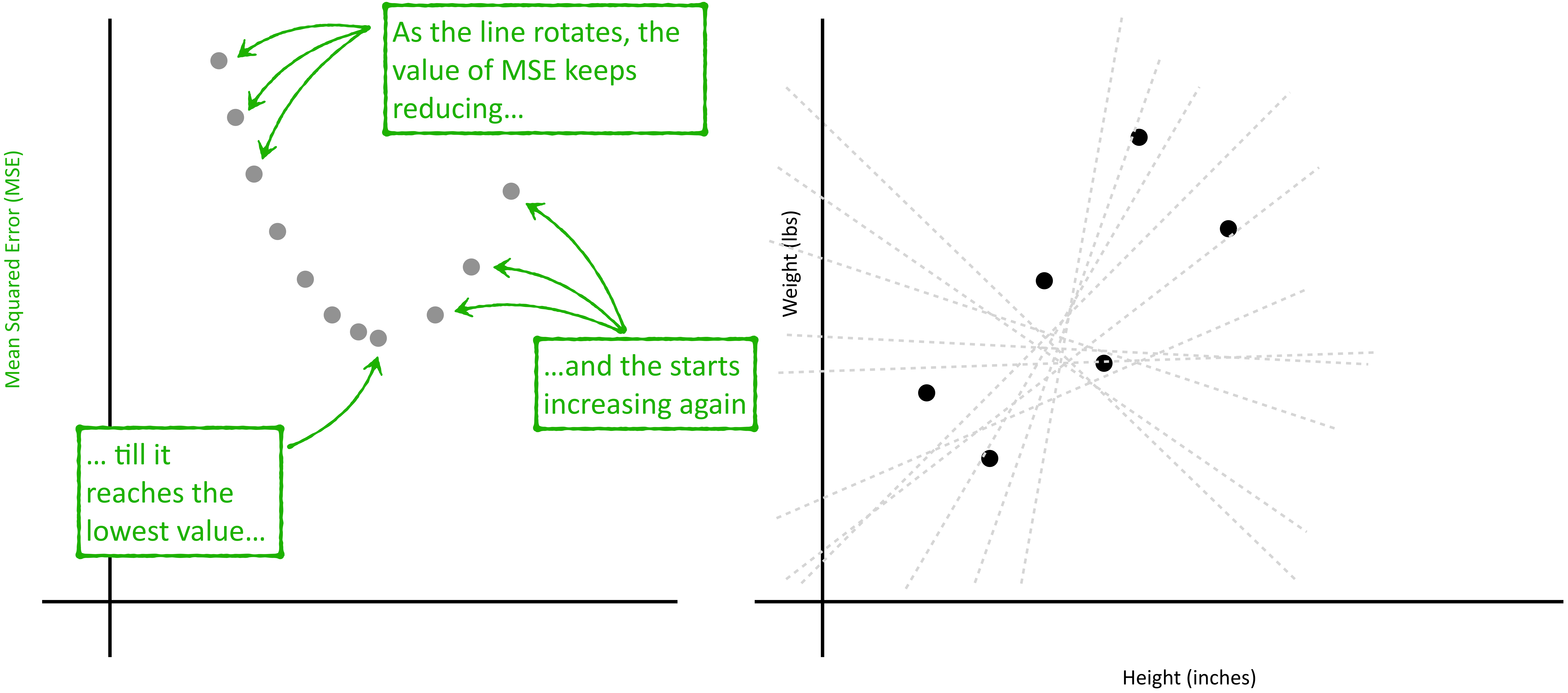
Simple Linear Regression

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1



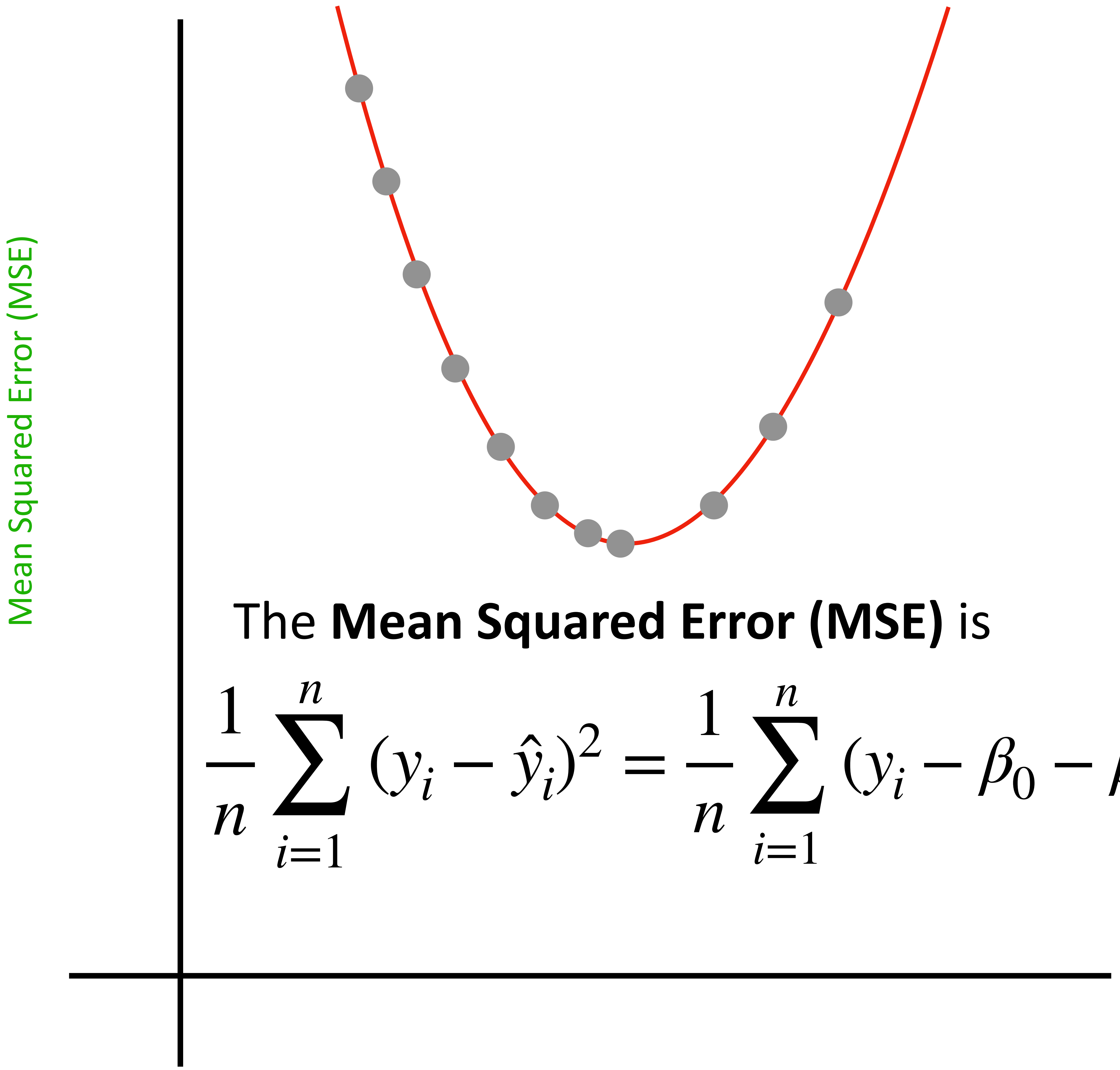
Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression



Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

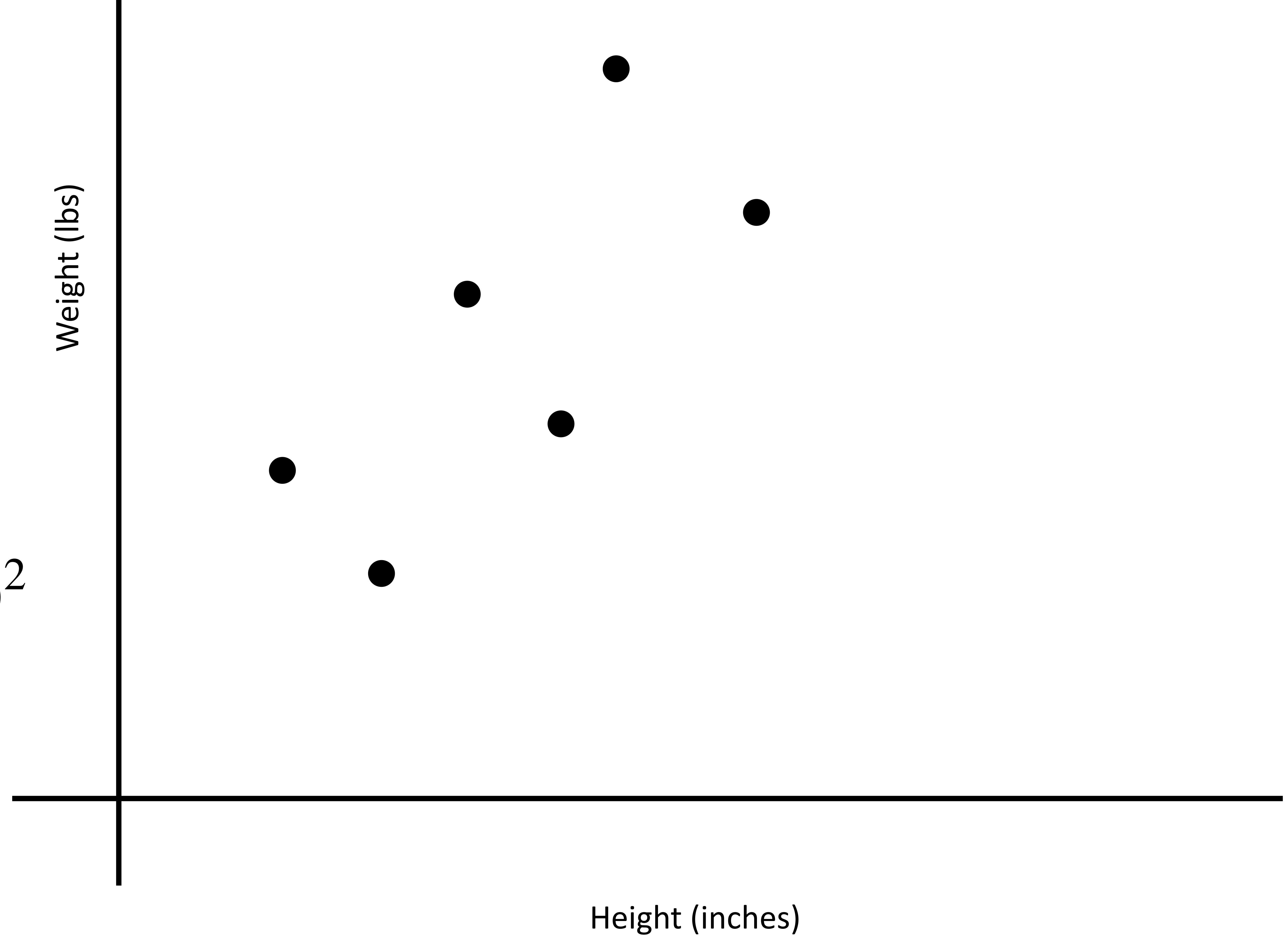
Simple Linear Regression



The **Mean Squared Error (MSE)** is

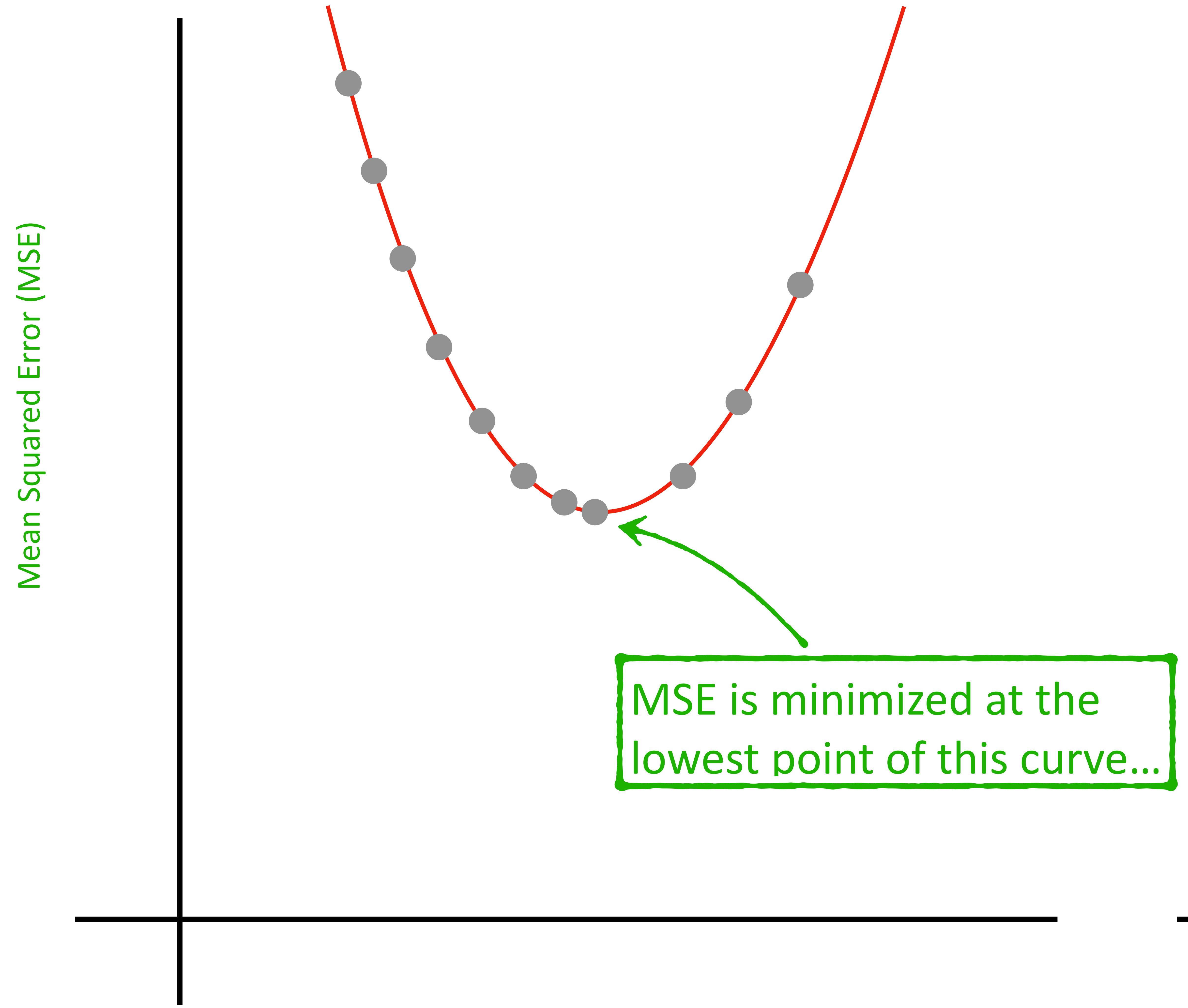
$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

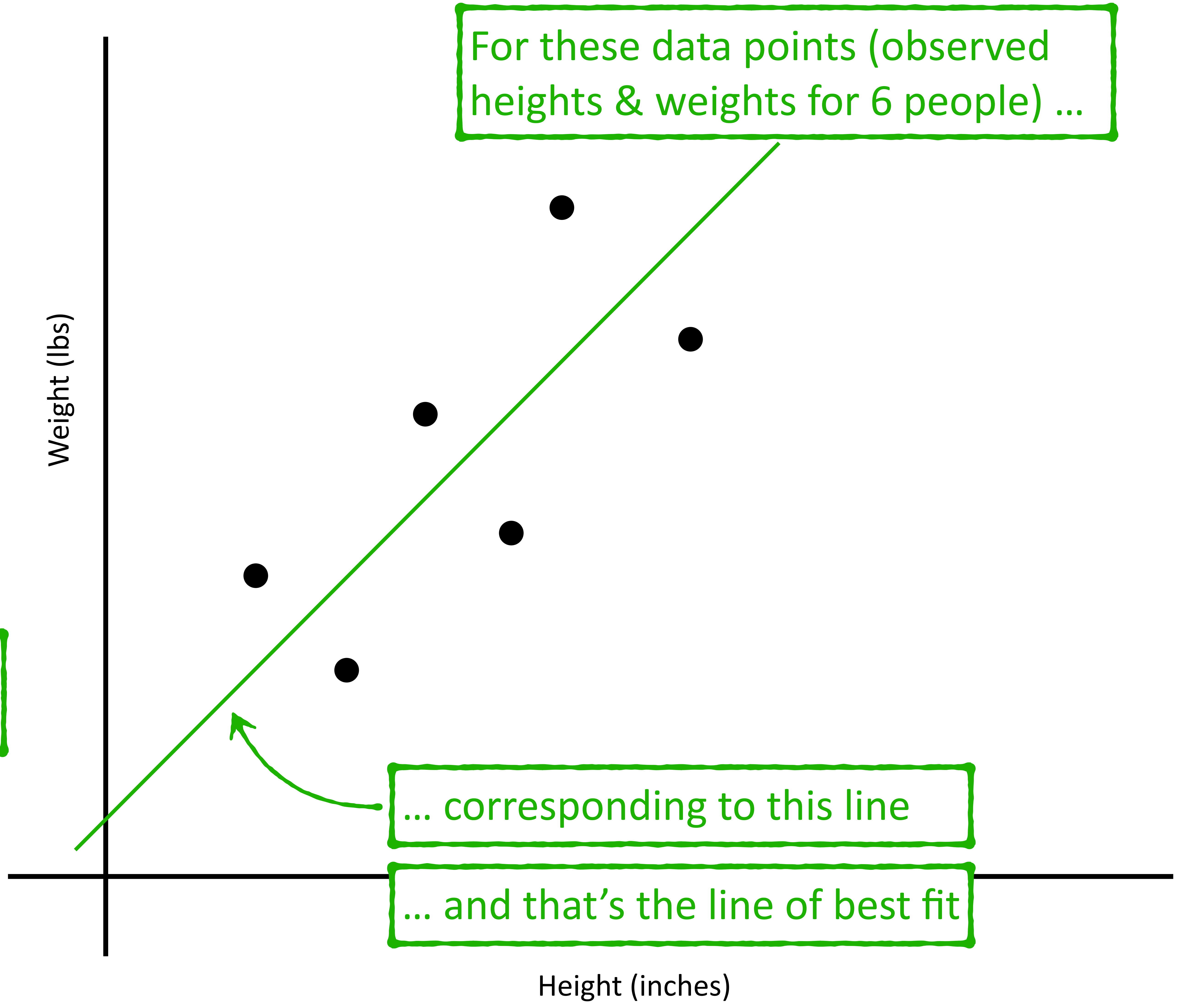
For these data points (observed heights & weights for 6 people) …



Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

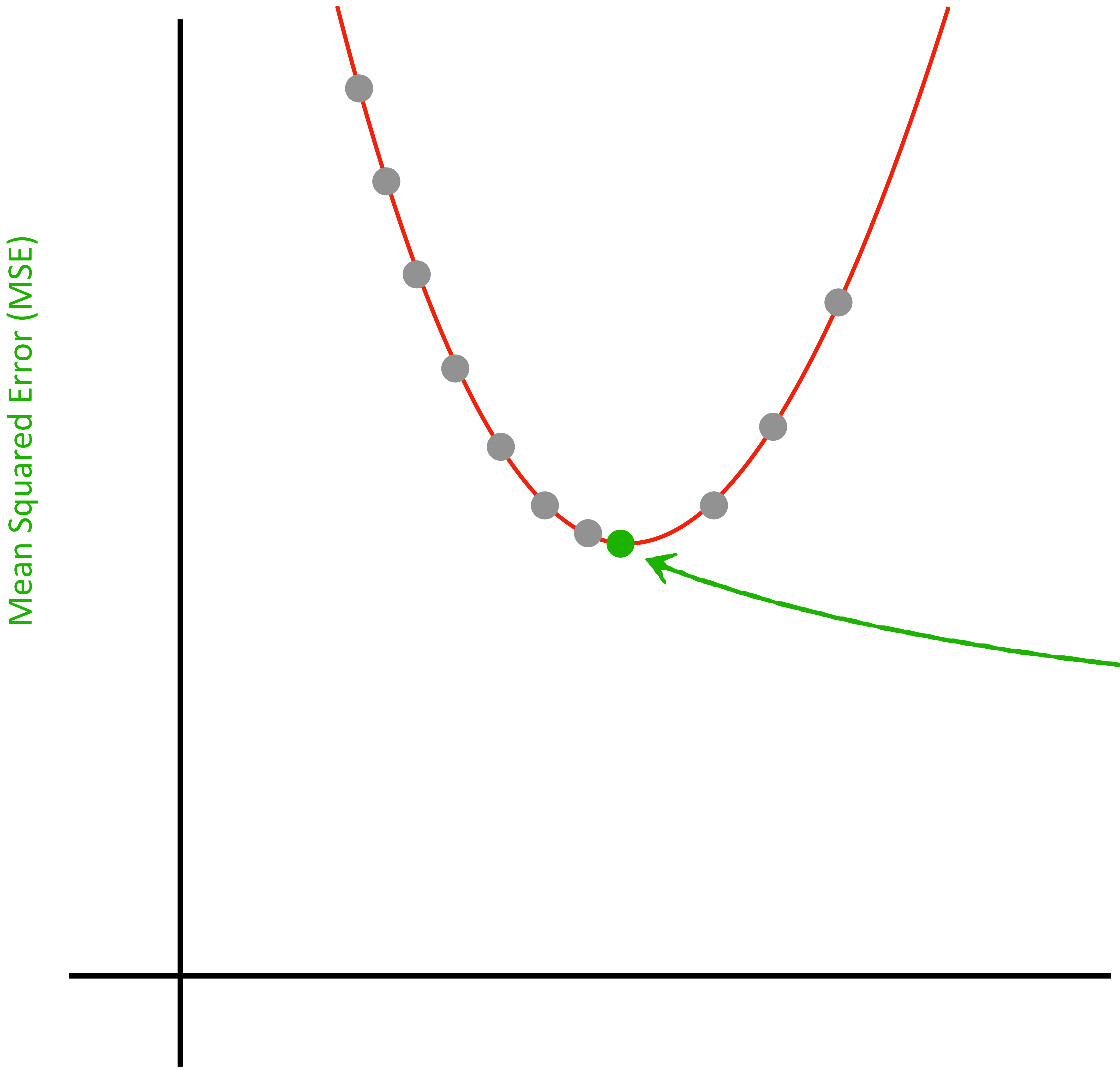
Simple Linear Regression





Simple Linear Regression

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1



MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(1)$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(2)$$

MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0
See Tutorial on Differential Calculus

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(1)$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(2)$$

Solving both equations for β_0 and β_1 we get…

MSE is minimized when the first derivative w.r.t β_0 and β_1 equals 0

See Tutorial on Differential Calculus

Lets calculate the **Mean Squared Error (MSE)** for various values of β_0 and β_1

Simple Linear Regression

The **Mean Squared Error (MSE)** is

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial \beta_0}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(1)$$

$$\frac{\partial}{\partial \beta_1}\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \cdots \; eq(2)$$

Solving both equations for β_0 and β_1 we get…

$$\beta_0 = \frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n}$$

$$\beta_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

This is known as the **Closed Form Solution** for Simple Linear Regression

For the details on how the two equations are solved see
Proof of the Closed Form Solution

Related Tutorials & Textbooks

Simple Linear Regression - Proof of the Closed Form Solution

A detailed proof of the the closed form solution of simple linear regression introduced above. This proof walks through solving two partial differential equations to compute the values of the two parameters.

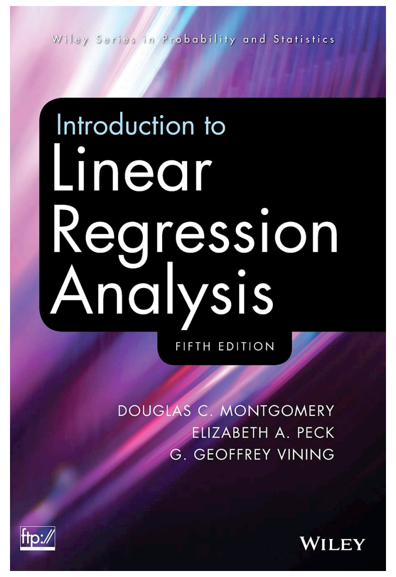
Multiple Regression

Multiple regression extends the two dimensional linear model introduced in Simple Linear Regression to k + 1 dimensions with one dependent variable, k independent variables and k+1 parameters.

Gradient Descent for Simple Linear Regression

An introduction to the Gradient Descent algorithm and a deep dive on how it can be used to optimize the two parameters β_0 and β_1 for Simple Linear Regression.

Recommended Textbooks



Introduction to Linear Regression Analysis

by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining

For a complete list of tutorials see:
https://arrsingh.com/ai-tutorials